Achievement motivation amongst Rwandan secondary students. An empirical investigation of its internal and external nomological network.

Dominik Bulla

St Cross College

DPhil in Education

Department of Education

University of Oxford

Supervisors

Professor Lars-Erik Malmberg

Professor Steve Strand

Acknowledgements

I would like to express my deep gratitude for all the support provided to me by my two supervisors, Professor Steve Strand and Professor Lars Malmberg. They have given me a constant support in person as well as online. I deeply appreciate all the time they have spent helping me complete this project. Thank you.

I would also like to thank the UK Economic and Social Research Council (ESRC) and St Cross College for providing me with the funding required to complete this DPhil. Without their financial support, I would not have been able to complete this research.

I dedicate this work to both my children Keza and Karla. I know at some point you will both be able to read these words.

Abstract

Achievement motivation refers to those cognitive-affective factors that are assumed to predict students' behavioural engagement and academic achievement. Within the Western-borne literature, achievement motivation has been conceptualised as a multidimensional, multileveled concept that holds educational relevance. It is of educational relevance because it is assumed to predict behavioural engagement and academic performance. It is multidimensional because, according to expectancy-value theory, behavioural engagement is driven by one's success expectations ("Can I do it?"), value perceptions ("Is it useful?"), and cost evaluations ("What does it require from me?"). It is multileveled since achievement motivation does not only determine behavioural engagement in specific contexts but also engagement and academic performance at the school subject-specific level. Given the current learning crises within many African countries, it is remarkable that achievement motivation has not been systematically studied within African contexts. Through a series of studies, this DPhil project examined the multidimensional and multileveled structure of achievement motivation as well as investigated its relations with classroom activities as well as behavioural engagement. Within **the first study** (n = 171, average age = 15.27; SD = 1.77), experiential descriptions of confidence and motivation were collected through open-ended questionnaires. The qualitative content analysis revealed that the expectancy-value model of achievement motivation adequately summarised students' experience of confidence and motivation. The themes identified were subsequently translated into self-report scales to measures achievement motivation quantitatively. Within **the second study** (n = 769; average age = 15.01; SD = 1.57), the generalised internal/external frame of reference (GI/E) model was applied to explain expectancy, utility value, and cost beliefs across chemistry, English, Kinyarwanda, and mathematics through prior school grades. Multigroup structural equation

modelling (SEM) of the cross-sectional data revealed that school type may moderate GI/E model relations. Amongst boarding schools, prior achievement predicted expectancy beliefs whereas amongst basic-education schools it predicted cost beliefs. In Rwanda, boarding schools are considered schools of excellence whereas basic-education schools are considered to be for those that fail academically. Within **the third study** (n = 446; average age = 15.33; SD = 1.59), the GI/E model was applied to average competence and cost states inferred through a latent trait/state model. Self-reported motivational states were lesson-specific and collected after English, Kinyarwanda, and mathematics lessons for 13 consecutive school days. The model hardly explained average states. Prior achievement only predicted expectancy traits within mathematics. Within **the fourth study** (n = 868; average age = 15.47; SD = 1.59), the effects of classroom activities (i.e., student-centred versus teacher-centred) on lesson-specific motivational states were examined. Latent trait/state modelling suggested motivational effects to vary across subjects. In mathematics, student-centred activities exerted mainly positive effects. In Kinyarwanda, teacher-dominant activities exerted mainly negative effects. In English, teacher explanations exerted especially positive effects. The effects of teacher explanations appeared to be mediated by a measure on perceived novelty. Within **the fifth study** (n = 188; average age = 15.1; SD = 1.62), the effects of both classroom activities and lesson-specific motivational states on situation-specific behavioural engagement, determined through external ratings, were examined. Three-level modelling revealed student-centred classroom activities did and motivational self-reports did not predict engagement. The latter may be due to the way the measures were operationalised. The project results are presented in the form of five manuscripts submitted to peer-reviewed journals. Overall, the present DPhil project revealed that the manifestation of achievement motivation is multidimensional but less nuanced, multileveled, but less interpersonally variable compared to the way it has been observed especially amongst

Western populations. There is evidence that achievement, too, holds educational relevance within the Rwandan context. However, additional research is necessary to further explore relations between achievement motivation and particularly behavioural engagement.

Contents

List of Figures

List of Tables

Introduction

1.1. The thematic focus of the present DPhil project

Motivational research within educational psychology is not only about mapping out the psychological factors that drive students’ engagement and learning in class. It has increasingly been focussing on translating respective insights into interventions that help improve students’ levels of engagement and learning at school. Evidence so far suggests that motivational interventions are not only effective but also cost efficient. As such, motivational interventions should be of interest to low-income countries that commonly struggle with the double whammy of having to address widespread low academic attainment amongst students within education while facing resource constraints within educational budgets. Against this background, it is unfortunate that within mainstream motivational research the empirical work has largely been centred on populations from the so-called Western hemisphere (e.g., Cornelius-White, 2007; Kriegbaum et al., 2018; Lei et al., 2018; Möller et al., 2009, 2020; Orth et al., 2020; Roorda et al., 2011). By contrast, Africa appears to be largely underrepresented as a research context. This is unfortunate because effective motivational programming of educational interventions requires a sound understand of the motivational processes at work amongst those students whose levels of engagement and learning one hopes to improve. Without a sound scientific understanding of motivational processes within African classrooms, motivational interventions cannot be adjusted to the local context. This DPhil project was about motivational processes at play amongst secondary students enrolled at public schools in Rwanda, East Africa. It thus helped not only to better understand to what extent current postulates of mainstream motivational theory, developed in the predominantly Western world, also generalize to African populations. It also constituted a first, systematic exploration of the links between achievement motivation reported by African school students and their levels of engagement in class and academic achievement across school subjects. It is intended that especially the latter may help to kick start evaluation research around the use and usefulness of motivational interventions within low-income countries of Africa.

To structure the research within the context of this DPhil project, the nomological net associated with achievement motivation was explored amongst Rwandan secondary students. Being the "why of behaviour" (Deci & Ryan, 1985, p. 3), motivation people experience is assumed to have explanatory power vis-à-vis those actions they take within different life domains. Within the educational domain, this nexus is generally discussed under the heading of achievement motivation. Achievement motivation as a concept referring to those psychological processes that translate a person's cognitive-affective (self-) evaluations (i.e., motivational factors) that emerge within school contexts into various education-related behavioural outcomes such as engagement (e.g., Connor et al., 2009; Fisher et al., 2014; Gobel, 2008; Godwin et al., 2021; Guillaume & Khachikian, 2011; Kupiainen et al., 2014; Moffett & Morrison, 2020)., academic achievement (e.g., Hailikari et al., 2008; e.g., Kriegbaum et al., 2018; Lavrijsen et al., 2021; Steinmayr & Spinath, 2009), or career choices (e.g., Lauermann et al., 2017). Within the present DPhil project, the study of achievement motivation amongst Rwandan students was approached through examining key elements of its "nomological net" (Cronbach & Meehl, 1955, p. 290). Within the scale-development literature (e.g., Benson, 1998; L. A. Clark & Watson, 2019), the nomological net is a description of the hypothesized internal as well as external manifestation associated with the theoretical representation of a concept of interest. A concept's internal manifestation refers to its latent structure whereas its external manifestation refers to how the different elements of the inner structure relate to those external variables that are considered to be of theoretical relevance for the concept concerned (Byrne, 1984). Thus, the nomological net provides a framework through which to organise the empirical investigation of a concept's internal as well as external structure it is hypothesised to exhibit (e.g., Benson, 1998; Byrne, 1984; Cronbach & Meehl, 1955; Rauthmann et al., 2020). Within the present DPhil project, the nomological net is used as an organisation devise to test whether or not the internal and external structure of achievement motivation manifest within Rwandan school context in the way currently discussed in the (Western-borne) educational literature.

Within the (Western-borne) educational literature, the internal structure of achievement motivation is generally assumed to be multidimensional and increasingly multileveled. Its multidimensional conceptualisation refers to the notion that student behaviours within performance contexts are shaped by more than just one cognitive-affective determinant. Within the context of this DPhil project, the multidimensionality of achievement motivation is organised through the framework of expectancy-value theory (e.g., Eccles & Wigfield, 2020). According to expectancy-value theory, the cognitive-affective determinants (i.e., motivational factors) are grouped into (self-) perceptions about how well one will perform on those activities that make up a performance context (i.e., expectancy evaluations), whether or not those activities are enjoyable, useful, and important (i.e., value evaluations), and how taxing these activities are (i.e., cost evaluations). Especially the value and cost constructs[1] has been increasingly differentiated out into different sub-categories (e.g., Flake et al., 2015; Gaspard et al., 2015). Within the present DPhil project, it was examined to what

[1] Within the context of this DPhil project, achievement motivation is considered a concept whereas expectancy, value, and cost (self-)evaluations are considered constructs. A concept is a "an idea that unites phenomena […] under a single term" (Bollen, 1989, p. 180). Within this DPhil, achievement motivation is considered to be the concept that unites the constructs of expectancy, value, and cost (self-)evaluations under a single, overarching term. Latent constructs are phenomena that cannot be observed directly (McCoach et al., 2014). Neither achievement motivation as a whole nor expectancy, value, and cost (self-) evaluations are directly observable. Thus, they are all latent constructs. To highlight that achievement motivation is not only a latent phenomenon but a phenomenon that comprises of different latent constructs, achievement motivation is described as a concept not construct.

degree the hypothesized multidimensional structure of expectancy, value, and cost (self-)evaluations describe motivational experiences of Rwandan students.

Achievement motivation is considered a multilevel phenomenon in the sense that is assumed to operate at different levels of specificity. A level of specificity can be loosely defined as how general a motivational experience is (Fishbein & Ajzen, 2010). At the situational level, motivation describes a person's (self-)experiences that emerge within a concrete context at a specific moment in time (i.e., motivational states). These (self-)evaluations may just apply to this very specific moment. Higher levels of specificities refer to (self-)evaluations that are typical for contexts that are increasingly more general (Fishbein & Ajzen, 2010). At the school subject-specific level, for example, motivation refers to those (self-) experiences that are typical within a given school subject. These (self-) experiences may not apply to each and every specific learning context within the school subject, but they are the most typical ones across all learning contexts within that subject. Because it represents typical experiences, school subject-specific motivation is assumed to be largely time independent (i.e., stable; e.g., Boekaerts, 1996; Bong & Skaalvik, 2003; Shavelson et al., 1976) whereas motivational states are considered time dependent. Motivational states are time dependent in the sense that they are shaped by those contextual factors that characterise a specific learning situation. Due to the dependency on time-bound contextual factors, concrete motivational states may thus divert from those (self-)experiences students typically associate with a school subject. Thus, motivational states can vary substantially even within short time intervals (i.e., short-term intraindividual variability; e.g., M. Martin & Hofer, 2004). The extent of intrapersonal variability of cognitive-affective states also provides an indication of the hypothesized multilevel structure of achievement motivation. It is only to the extent that motivational states show at least some intrapersonal variability that they are distinct from those stable motivational tendencies that are specific for a school subject or person. Within the present DPhil project, the hypothesis of intrapersonal variability of motivational states amongst Rwandan students was tested through an intensive longitudinal research design (e.g., Bolger & Laurenceau, 2013). The present DPhil project might have been the first research undertaking that applied an intensive longitudinal design within the Sub-Saharan school context.

Achievement motivation varies across levels of specific not only in terms of temporal stability but also in terms of the correlates it is associated with (e.g., Swann et al., 2007; Vallerand, 1997). At the school subject-specific level, motivation is assumed to predict (e.g., Lavrijsen et al., 2021) and be predicted (e.g., Gaspard et al., 2018) by school grades. In concrete learning contexts, achievement motivation is assumed to be predicted by classroom activities (e.g., Jansen, Scherer, et al., 2015) and to predict engagement (e.g., Dietrich et al., 2017; Olivier et al., 2020; Salmela-Aro et al., 2021). Within the present DPhil project, the links between expectancy, value, and cost (self-)perceptions on the one side, and school subject-specific school grades, lesson-specific classroom activities, and situation-specific on-task behaviour on the other side were examined.

Achievement motivation describes person-internal factors and thus is commonly measured through self-reports (Appleton et al., 2006). At the situational level, researchers may ask respondents about the concrete motivational experiences they have right now in this moment (e.g., Bolger & Laurenceau, 2013). To measure achievement motivation at a higher level of generality (e.g., the school subject-specific level), researchers face different options to derive at those experiences that a typical for an individual within a general context. In line with the option most used in educational research, researchers ask students to indicate what they believe their typical motivational experiences are, for example, within a given school subject (i.e., school subject-specific self-reports). Another option implies determining the trend within self-reported motivational states that have been repeatedly taken from students within a given time interval. For example, to obtain a measure of typical experience made, researchers can determine the mean of self-reports taken (Fleeson & Gallagher, 2009). The former approach is more economical than the latter approach. To obtain a measure of typical experiences, one does not need to ask respondents repeatedly but just once (e.g., Pekrun & Buehler, 2014). The former approach also seems to have some educational value. School subject-specific self-reports are found to be related to external variables such as school grades (e.g., Gaspard et al., 2018). However, students have also been found to be bad at determining their typical experiences within a given domain (e.g., Götz et al., 2013). Aggregating psychological states across time into one overall rating is cognitively taxing and difficult. Hence, school subject-specific self-reports may be best interpreted as beliefs about one's own typical states (Robinson & Clore, 2002; Roos et al., 2015). In comparison, average motivational states may constitute valid representations of one's typical motivational experiences. However, the extent to which they relate to external variables has hardly been examined within educational contexts. This in turn is important because it is only through the relations with external variables that a construct obtains "meaning" (Benson, 1998, p. 11; Crocker & Algina, 2006). Within the present DPhil project, the educational meaning of both school subject-specific motivational self-reports and average motivational states was studied through relating them to prior school grades. The empirical work in this respect was guided by the generalised internal/external frame of reference (GI/E) model (e.g., Möller et al., 2016), which provides a set of hypotheses of how prior academic achievement relates to subsequent motivational levels within and across school grades. To sum up, the focus of this DPhil research was about the extent to which achievement motivation expressed by lower-secondary students in Rwanda is multidimensional, multileveled, and related to school grades, classroom activities, and engagement. The rationale for choosing Rwanda as the DPhil research context is twofold. Rwanda, a landlocked low-income country in East Africa, was chosen both on theory-related and policy-related grounds. They outlined in the next two sections of this introduction.

1.2. Rwanda as the research context: the theoretical rationale

The way achievement motivation manifests itself has increasingly been recognized as culturally dependent (e.g., Eccles & Wigfield, 2020). It describes the notion that the way the internal and external structures of achievement motivation manifest themselves varies across cultures. Thus, to distinguish those aspects of achievement motivation that are inherent to all humans from those aspects that are inherent to some cultural contexts requires motivational research to be based on a large and especially diverse database of study samples (Henrich et al., 2010; R. B. King et al., 2018; Zusho & Clayton, 2011). Only if reviews have access to data from a diverse set of different cultural contexts can they tease out those universal aspects from the culturally specific aspects (e.g., M.-T. Wang et al., 2019). Against this background, it matters if entire regions are underrepresented within motivational research. Geographic distributions of studies used within meta-analyses of various aspects of achievement motivation (e.g., Cornelius-White, 2007; Kriegbaum et al., 2018; Lei et al., 2018; Möller et al., 2009, 2020; Orth et al., 2020; Roorda et al., 2011) suggest motivational research is reliant especially on populations from the Western, educated, industrialized, rich, educated, and developed (i.e. "WEIRD"; Henrich et al., 2010, p. 61) parts of the world. By contrast, African populations are hardly included into research on achievement motivation. At the very minimum, it implies that it is not known to what extent current motivational theories also generalise to African contexts[2]. Through carrying out data collection in Rwanda, East Africa,

[2] Given the vast diversity in Africa in terms of climate, culture, religion, socio-economic conditions, and political structure, it would not be justified to speak of one African *context*. Instead, the continent of Africa is best understood as a multitude of different *contexts* that may vary in terms of their similarity.

the present DPhil project thus contributed to a diversification of the current database of study samples within motivational research. This contribution was not only empirical (since it added empirical data on achievement from a cultural context not previously studied) but also theoretical. It was theoretical in that the primary data collected were used to compare the observed manifestation of achievement motivation amongst Rwandan secondary students with the conceptualisations of achievement motivation currently discussed within the (Western-borne) educational literature. That way, this DPhil project contributed to identify what Berry (1989) calls the "universals" (p. 728) amongst the "imposed etics" (p. 727). Applied to the case of achievement motivation, imposed etic describes those aspects of the nomological net that have been observed within one or some culture. It has not been established yet whether or not these aspects hold true across a number or even all cultures. Hence, imposed etics constitute a set of assumptions about the way a culturally dependent concept manifests itself within different cultural contexts that can be tested. To the extent that respective aspects have been found in other contexts as well, they do not have to be considered imposed anymore. Hence, given that achievement motivation is culturally dependent, the development of motivational theory that generalises across different cultures requires its examination within the different cultures concerned. This has also been highlighted by the universalist research perspective[3] on achievement motivation (e.g., R. B.

[3] With regards to the study of psychological constructs within cultural contexts, one distinguishes between the etic, emic, and universalist research. Within etic research, it is the scientist who defines and operationalises psychological constructs of interest. Within emic research, one attempts to discover the actual meaning and relevance of psychological phenomena within the culture it appears by adopting a purely local perspective (Berry et al., 2002; Chao & Lambert, 2013; F. M. Cheung et al., 2011). Accordingly, psychological constructs are not derived from any existing, potentially outside theory at all. Instead, one aims at unpacking meaning of cultural concepts using local knowledge and interpretations uncovered through lexical approaches or through open-ended, semi structured observation and interview techniques involving those that live within the culture studied (Berry et al., 2002; Chao & Lambert, 2013; F. M. Cheung et al., 2011). The universalist research approach, as outlined in the main text, constitutes a middle ground between etic and emic research. It assumes that key concepts can be derived from general theory but their local manifestation needs to be studied through the local lens (R. B. King et al., 2018; Zusho & Clayton, 2011).

King et al., 2018; Zusho & Clayton, 2011). According to Berry (1989), the development of more general theories requires the following sequence of steps. First, researchers should develop the theoretical concept and the associated nomological net within their own culture (i.e., culture A). In a second step, researchers should transport and test the theoretical framework in a different culture (i.e., culture B). Even though this is ideally done by researchers coming from the local contact under study, it does not require local researchers. Insofar as external researchers are willing to emerge themselves into the local context, they, too, can discover local realities. In a final step, empirical results are compared across culture A and B in order to identify those aspects that apply to both contexts (i.e., derived etic). The application of a theoretical framework developed in one's own culture to a different cultural context is justified as long as this framework is considered a starting point for subsequent comparisons and is allowed to be modified based on the comparative results (i.e., imposed etic; Berry, 1989, p. 726). In simple terms, the usage of imposed etic is justified because "all research must begin somewhere" (p. 726). Within the context of this DPhil project, the literature review in section two represents the first step within Berry's (1989) research programme. The conceptualisation of achievement motivation through the internal and external elements of the nomological that will be presented below is derived from the (Western-borne) educational literature and thus constitutes the imposed etic which is the applied to the case of lower-secondary students in Rwanda. Functionally speaking, the Western-borne educational literature on the nomological net within and around achievement motivation constitutes culture A whereas Rwanda, the geographical target of this DPhil research, constitutes culture B. Especially chapter five presents the results of the application of the Western-borne conceptualisation of achievement motivation, as the imposed etic, to the Rwandan context. The general discussion in chapter six is then the comparison of the concept of achievement motivation as well as its nomological net between Rwanda and the corresponding motivational literature (for a similar approach, see Clayton & Zusho, 2016). A comment is warranted here. Especially within the domain of expectancy-value theory, achievement motivation has been subject both to extensive research (e.g., Atkinson, 1958; Atkinson & Birch, 1970; Eccles et al., 1983; Eccles & Wigfield, 2020) as well as reflections on the importance of culture (e.g., Eccles et al., 2004). As the onset of this DPhil project, it was therefore not expected that the application of the nomological net within and around achievement motivation to the Rwandan context would result in a derived etic that substantially alter current conceptualisations of achievement motivation. It was rather expected that the comparison between Rwanda and the wider motivational literature would result in the identification of nuances that may exist within the manifestations of achievement motivation. Finally, it was not expected that research on the nomological net associated with achievement motivation in Rwanda would have been 'completed' through this DPhil project. If anything, it was hoped that this DPhil project could have helped kicking off the systematic study of achievement motivation within Rwanda, and the wider region of (East) Africa. As the next section highlights, this may have not only theoretical but also important practical implications.

1.3. Rwanda as the research context: the policy-related rationale

Even though Africa is underrepresented within mainstream motivational research (e.g., Cornelius-White, 2007; Kriegbaum et al., 2018; Lei et al., 2018; Möller et al., 2009, 2020; Orth et al., 2020; Roorda et al., 2011), research on motivation amongst African students has been carried out. Unlike motivational research within the region of the Economic Co-operation and Development (OECD), which predominantly seems to focus on achievement motivation within schools, research related to motivation in Africa has especially been concerned with the question of “Why do children go to school?” (Marshall, 2016, p. 424). It might have been reflective of the wider policy context especially of the past. In fact, for a long time, many African education systems had been suffering from low school enrolment and completion rates. In more recent years, however, many African countries have made major strides towards getting and keeping both boys and girls in school (e.g., Jones et al., 2014; Krishnaratne et al., 2013). Alas, the progress in terms of enrolment and attendance rates was hardly matched by progress in terms of learning outcomes. Within many African countries, the number of years students spent in school has often not been translated into adequate learning gains (Jones et al., 2014; Pritchett, 2013; Sandefur, 2018). Sandefur (2018), for example, compared mathematics-related test scores from 14 Sub-Saharan African countries with comparable test results from countries within the OECD. Linking scales through item response theory, he was able to combine mathematics results from the Southern and Eastern Africa Consortium for Measuring Education Quality (SACMEQ) assessment with the Trends in International Mathematics and Science Study (TIMSS). He found that the test scores obtained for the 14 Sub-Saharan African countries was approximately one-and-a-half standard deviations below the mean for the non-African TIMSS countries. It might be results like these that made observers conclude that “schooling ain't learning” (Pritchett, 2013), especially in many parts of Africa but also other developing countries. Measures to improve not only the quantity (e.g., completion rates) but also the quality (e.g., learning outcomes) of education have generally been grouped into supply- and demand-side interventions (Krishnaratne et al., 2013). Examples for the former include the provision of learning materials, school building, or teacher training. Examples for the latter include abolition of school fees. Most likely, methodologies currently discussed under the heading of education systems research (e.g., Magrath et al., 2019) are best positioned to better understand how the interplay of the different demand- and supply-side factors jointly determine school completion and learning outcomes (also see Borman & u.a, 2003). However, ignoring the motivational effects of supply- and demand-side interventions within education "will undermine system effectiveness" (p. 604), as Lazowski and Hulleman (2016) argue. After all, it is only through the motivational avenue that school and classroom contextual factors shape students' levels of behavioural engagement, which in turn is the prerequisite of learning and thus learning gains (e.g., Skinner & Pitzer, 2012). In addition, achievement motivation does not only constitute a mediator of the effects of wider demand- and supply-side interventions. Achievement motivation also constitutes a field of educational interventions itself. For example, interventions that enhance students' perceived utility of content they study have been associated with positive effects on academic performance (e.g., Hulleman, Godes, et al., 2010). Lazowski and Hulleman (2016) carried out a meta-analysis of altogether 74 published and unpublished articles on the effects of motivation interventions. Across studies, they found an average mean effect of d = 0.52 (SE = 0.091) of motivation interventions on academic performance. This effect size is higher compared to the typical effects (d = 0.40; SE = 0.062) associated with the parent-level, teacher-level, school-level, and other student-level determinants to learning (Hattie, 2009). Moreover, in contrast to comprehensive school reforms, which are extremely resource-intensive (e.g., Borman & u.a, 2003), motivation interventions are described as rather cost effective (e.g., Yeager & Walton, 2011). In fact, Lazowski and Hulleman (2016) argue that most of the treatments covered within their meta-analysis of motivational interventions would be more or less cost-free. For example, the above-mentioned relevance intervention evaluated by Hulleman, Godes, et al. (2010) utlity percpetions were manipulated through a writing task that required from students to explain in what way the content they were about to study was useful for them. The cost-efficiency and effectiveness of motivation interventions should make them a promising option to improve educational outcomes for all education systems. Within the OECD world, they are indeed becoming more relevant (e.g., Lazowski & Hulleman, 2016; Yeager & Walton, 2011). By contrast, cognitive-affective interventions to manipulate students' experience of achievement motivation are not common at all within African contexts (e.g., Conn, 2017). One explanation could be that the lack of evidence on the nexus between achievement motivation and academic achievement (and other educational outcomes of relevance) may prevent cognitive-affective interventions to attract attention by education policymakers and practitioners. Hence, systematic research that provides evidence on the links between achievement motivation and academic performance within African contexts may help to build a case for incorporating motivation interventions into education programming within African countries. This was the policy rationale for choosing an African country as the context for the present DPhil project. Rwanda was chosen because it has been one of the best-performing countries in terms of providing education for all (UNESCO, 2014). By 2012, primary education enrolment was almost universal (Williams, 2019) while lower-secondary net enrolment was about 32 per cent in 2019 (MINEDUC, 2019). Despite the progress towards universal primary and increasingly secondary education, Rwanda, too, has been struggling with poor academic achievement of its students (Burdett & James, 2018). From an educational point, Rwanda has been a well performing but not atypical low-income country within Africa. Finally, Rwanda's governance has been described as "developmental" (Williams, 2019, p. 648) in that political stability is maintained through governmental policies that move all Rwandans out of poverty. Government efforts thus focussed amongst others on transforming Rwanda from a subsistence-based agricultural economy into a knowledge-based economy (MINECOFIN, 2000). Thus, the political context of Rwanda may be conducive to the introduction of innovative initiatives such as motivational interventions as long as they have been shown to contribute to the country's drive towards becoming a knowledge-based society.

The purpose of the present DPhil project was not only to help build a case for motivational interventions within Africa's education systems. As much as motivational interventions have the potential to be both cost-efficient and effective, there is an important caveat to this optimism. Impactful motivational interventions require "a precise understanding of students' subjective experience in school" (Yeager & Walton, 2011, p. 283). This understanding helps to determine where, when, how, and for how long to intervene in students' motivational experiences in order to improve academic outcomes. These insights may not suffice to effectively scale motivational interventions across entire education systems, but they constitute an important start into researching what interventions should be tested and potentially scaled. Hence, in respect of research on achievement motivation helping to gain a better understanding of students' subjective experience in school, it can be seen as a contribution to improving learning outcomes. It is for this reason that it is argued that focus within motivational research within African contexts should increasingly be shifted away from why students go to school to what their subjective experiences in school are. The work around the internal nomological net of achievement motivation pursued within the present DPhil project must be seen in this light. Moreover, the research on its external relations especially with behavioural engagement and school grades was intended to highlight the policy relevance of achievement motivation vis-à-vis learning outcomes. Given that achievement motivation was found to be related to behavioural engagement and school grades, a particularly convincing case could have been made for further systematic research on achievement motivation within resource-constrained settings such as Rwanda, which struggles with learning outcomes. It was this intention that also drove the decision to carry out a DPhil project on achievement motivation not within a Western but an African context.

1.4. The overall structure of the thesis

This DPhil thesis follows the structure of an integrated DPhil thesis. Following this introduction, it continues with a review of the academic literature on achievement motivation in chapter two. The overall objective of the literature review is to outline key aspects of the nomological net within and around achievement motivation currently discussed within educational research. These aspects are derived from the predominantly Western-borne academic literature on achievement motivation and can be considered theoretical assumptions to be tested within the Rwandan school context. It was Western borne in the sense that the respective academic literature was influenced through empirical (and theoretical) work based on Western sample. The reliance on the Western-borne literature was necessary because, so it appears, research on achievement motivation within African school contexts has hardly been reported if carried out at all (see previous section). The key elements of the nomological net within and around achievement motivation were translated into altogether five sets of research questions, which guided the field work, data analysis, and reporting of this DPhil project. These research questions are established in chapter three and concern achievement motivation's internal structure (e.g., the level of multidimensionality) as well as its relations with external variables such as prior school grades and engagement. The fourth chapter is about the overall methodology chosen to implement this DPhil project in Rwanda. It outlines the organisation of the data collection in Rwanda and the analysis approach taken to answer the research questions. The fifth chapter presents the results obtained in response to the different sets of research questions. In line with the format of an integrated thesis, the results are presented in form of five manuscripts written for five peer-reviewed journals. As manuscripts, these articles have partly been submitted to peer-reviewed journals. The thesis concludes with a general discussion that cuts across the findings associated with each of the five manuscripts.

2. Literature review

The following chapter provides an overview of the academic literature on achievement motivation. To structure the discussion, the theoretical framework of a nomological net (Cronbach & Meehl, 1955) is applied to the concept of achievement motivation. Within the construct-validity research, the nomological net provides an outline of both the internal structure of a concept under study as well as its hypothesised relations with external variables (Byrne, 1984). Within section 2.1, the internal structure of achievement motivation is outlined. The section starts by providing a definition of the concept of achievement motivation. It then introduces achievement motivation as both a multidimensional and multileveled concept. The latter refers to achievement motivation being measured at different levels of specificity (Fishbein & Ajzen, 2010) such as the lesson-specific or school subject-specific level. Multidimensionality and multileveledness are considered two key aspects of the internal structure of achievement motivation. Section 2.1. also provides a discussion on how achievement motivation is measured empirically, which constitutes a key element of a concept's nomological net (Cronbach & Meehl, 1955). At this point within section 2.1, school subject-specific self-reports as well as average motivational states as two different operationalizations of motivational dimensions at the school subject-specific level are discussed.

Section 2.2 is about the external nomological network. Here, relations between external variables and the different dimensions of achievement motivation across the different levels of specificity are discussed. The study of the internal structure of achievement motivation within the Rwandan context may be of theoretical interest. However, it is only through its relations with external variables that the concept of achievement motivation gains "meaning" (Benson, 1998, p. 11; Crocker & Algina, 2006). Of practical importance may especially the relations with external variables achievement motivation is assumed to predict. Only to the extent that achievement motivation is shown to predict educational outcomes of importance it may obtain practical relevance. Thus, to build a case for the study of achievement motivation not only within the OECD world but also within low-income countries such as Rwanda, one needs to demonstrate that achievement motivation predicts educational outcomes in low-income settings such as Rwanda. In terms of external educational variables of interest, section 2.2 introduces academic achievement, behavioural engagement, and classroom activities.

In section 2.3, achievement motivation is discussed as a “cultural” variable. It describes how culture shapes both the internal as well as external elements of the nomological net within and around achievement motivation. The implications are obvious. Since culture moderates internal and external relations associated with a psychological concept, empirical findings from one cultural context cannot easily be transferred to another cultural context. Because empirical patterns involving achievement motivation may vary across cultures, the development of motivational theories requires empirical work that is not limited to just one region of the world (e.g., the Western hemisphere) but that is culturally as inclusive and diverse as possible. It is for this reason that the dearth of research on achievement motivation within the African context is not only of practical but also of theoretical relevance. Section 2.4. summaries the evidence available on achievement motivation in Africa. It highlights that the different elements of the nomological net within and around achievement motivation have hardly been studied systematically within the Sub-Saharan context. Section 2.5. introduces the reader to the Rwandan country context, which also concludes the literature review.

2.1. The internal nomological network of achievement motivation

2.1.1. Defining achievement motivation as a multidimensional concept

Motivation represents “the process whereby goal-directed activities are instigated and sustained” (Schunk et al., 2013, p. 5). As a concept, motivation is used not only to study what activities (and behaviours) individuals go for (“choice”) and for how long (“persistence”) but also the level of involvement they actually exhibit within each of these activities (“effort”; Dörnyei, 2000, p. 520; Wigfield et al., 2006). Even though choice, persistence, and effort are also shaped through physiological drives and tensions (Deci, 1992), social-cognitive theories describe motivation as a process that is predominantly psychological in nature (Schunk et al., 2013). Accordingly, motivation is seen as the interaction of person-internal experiences such as emotions (Pekrun & Perry, 2014), goals (Elliot & Hulleman, 2017), psychological needs (Ryan & Deci, 2017), perceived ability (Bong & Skaalvik, 2003) value beliefs (Gaspard et al., 2015), or success attributions (Weiner, 1985) that jointly influence a person’s behaviours and actions. It is for this reason that the internal structure of achievement motivation needs to be conceptualised multidimensionally.

Achievement motivation applies the study of motivational processes to the educational domain. It refers to the psychological processes that translate a person’s cognitive evaluations and affective experiences (i.e., motivational factors) into those actions that are eventually exhibited within education-relevant learning and performance challenges (Wigfield et al., 2006, 2015). Theories on what person-internal factors achievement-motivational processes involve are manifold (Hattie, 2009; e.g., Schunk, 2013). In terms of numbers, they have grown to the point that “we need to critically consider whether the field truly needs so many distinct theories and constructs” (Anderman, 2020). Eccles’ et al. (1983)

expectancy-value theory constitutes probably one of the longest standing and most eminent models of achievement motivation. Based on earlier models (e.g., Atkinson, 1957), it was developed to explain students' academic choices, along with their persistence and performance within education (Eccles et al., 1983). These achievement behaviours are assumed to be the product of expectancy, value, and cost (self-)evaluations. These constructs are discussed below. A person's expectancies were initially assumed to be shaped by self-perceptions about how well they would perform on an upcoming academic challenge (i.e., "success expectations"), which in turn were assumed to be shaped in particular by self-evaluations about how well they performed during past academic challenges (i.e., "self-concept of ability"; Eccles et al., 1983; Wigfield & Eccles, 2000). Factor analytical work, however, suggested both factors to be inseparable (e.g., Eccles, Wigfield, et al., 1993) and thus success expectations and ability beliefs have increasingly been used interchangeably (e.g., Wigfield & Gladstone, 2019).

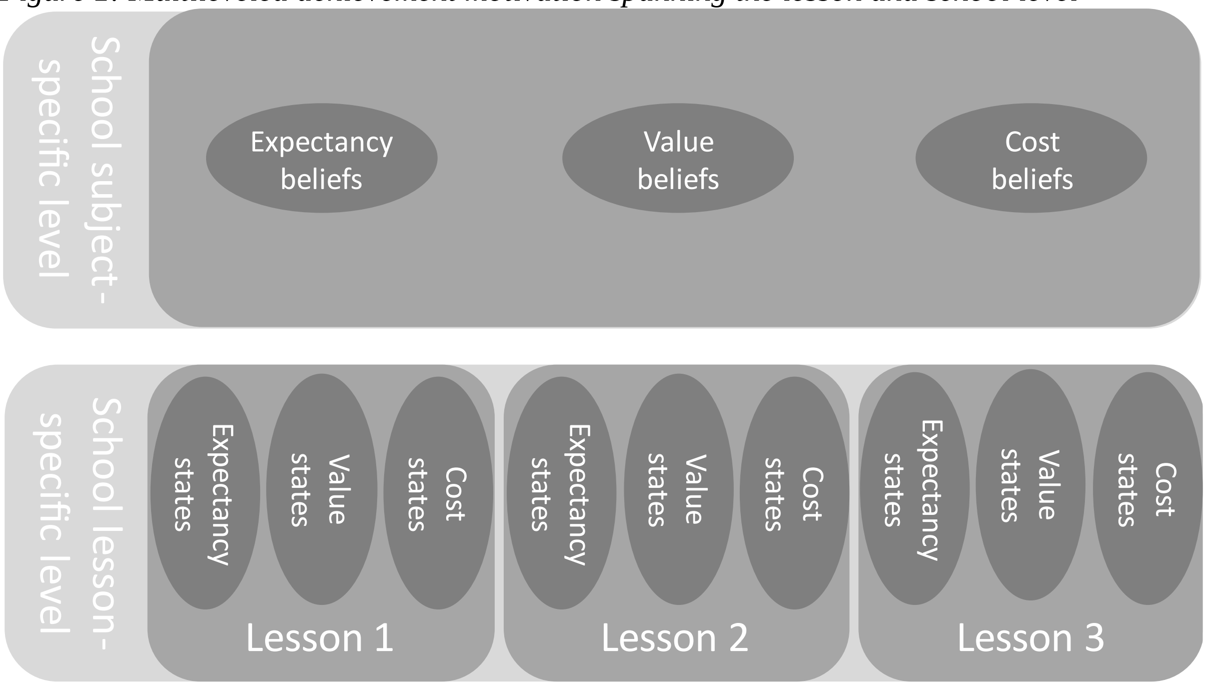
Eccles et al. (1983) did not conceive expectancy-value theory but modified some of the postulates of its earlier versions (e.g., Atkinson, 1957), which made it fit for purpose within "modern" educational settings. One of the modifications concerned the multidimensional representation of the value component. As per modern expectancy-value theory, value perceptions are about the level of enjoyment academic behaviours bring about (i.e., "intrinsic value"), how useful the outcomes of these behaviours are for oneself (i.e., "utility value"), and how important the successful attainment of these outcomes is for one's own identity (i.e., "attainment value"; Eccles et al., 1983). Within the expectancy-value tradition, the value component has continued to be subject to revisions. Gaspard et al. (2015) reconceptualised attainment and utility value as two- and five-dimensional constructs respectively. For example, utility value was then about perceived utility for one's school career, for one's future profession, for one's current, social, and future life. Additionally, Hulleman and colleagues (e.g., 2015) proposed to restructure the top-level components of the expectancy-value framework. Initially, the value component also contained facets that represented a person's perceived investments required to be successful within a given academic challenge (i.e., "costs"; e.g., Eccles et al., 1983), an aspect that has often been neglected within motivational research (Flake et al., 2015). It was then proposed to consider costs not just a facet of value but to see it a key component alongside expectancies and value (i.e., "expectancy-value-cost model"; e.g., Barron & Hulleman, 2015). In fact, factor analysis suggests costs to be a multidimensional construct as well. It combines perceptions about how challenging (i.e., "effort costs") as well as psychologically taxing (i.e., "emotional cost") an academic objective is, about the extent to which other opportunities need to be given up in order to pursue that objective (i.e., "loss of valued alternatives"), and about how time and work other academic objectives one currently pursues require (i.e., "outside-effort costs"; Flake et al., 2015). However, it is not entirely clear to what extent these proposed modifications of the initial expectancy-value theory are indeed necessary. For example, Eccles and Wigfield (2020) themselves do not see any reasons to "add the C to EVT" (p. 6). It is also for this reason that within the context of this DPhil research project the term expectancy-value theory is used.

A key strength of expectancy-value theory as a way to represent the multidimensionality of achievement motivation is that it allows incorporating many of the aspects currently discussed in motivational literature in a parsimonious way (e.g., Wigfield & Cambria, 2010). This may be best illustrated through two different examples. A first example concerns the aforementioned self-concept beliefs. Within motivational theory, self-concept research has provided a wealth of insights into the internal and external nomological network of achievement motivation (e.g., Marsh, Seaton, et al., 2019). In fact, the study of one of its research outcomes, the internal/external frame-of-reference (I/E) model, forms an important component of this DPhil project. As already alluded to earlier, empirical work suggests a very close relatedness of both expectancy beliefs and self-concept beliefs to the point that they appear to be factor-analytically inseparable. This in turn implies that models concerning self-concept should also apply to expectancy beliefs. As a matter of fact, the I/E model has been applied to the value and cost dimension of expectancy-value theory as well (e.g., Gaspard et al., 2018). A second example concerns the theories on the affective side of motivation. Intrinsic value within expectancy-value theory has been compared with situational interest (Eccles & Wigfield, 2020), with the latter being defined as "focused attention and the affective reaction that is triggered in the moment by environmental stimuli, which may or may not last over time" (Hidi & Renninger, 2006, p. 113). In fact, the affective side of achievement motivation (i.e., achievement emotions) has been increasingly considered a key driver of students' achievement behaviours (e.g., Pekrun et al., 2007; Pekrun & Perry, 2014). Achievement emotions do not only appear in the form of positive (i.e., intrinsic value) but also in the form of negative affective experiences. The latter aspect is covered by emotional costs within the expectancy-value framework (Flake et al., 2015). Hence, expectancy-value theory may not only be a parsimonious starting point to explore achievement motivation amongst Rwandan students. Rather, it may also constitute a starting point for the study of achievement motivation that allows to easily branch out to other cognitive affective theories that one may want to subsequently apply to explain achievement behaviours amongst Rwandan students. Research undertakings, however, should not only contemplate what dimensions of motivation to include but also what level of specificity to focus on because achievement motivation is best understood as a multilevel construct as the next section highlights.

2.1.2. Achievement motivation as a multilevel concept

Motivational constructs such as expectancy beliefs or intrinsic value are increasingly represented in a multilevel way (e.g., Götz et al., 2006; Shavelson et al., 1976; Vallerand, 1997). It implies the notion that these constructs vary not only between individuals at any given point in time (interpersonal differences) but also between time points for any given individual (intrapersonal differences). Within the literature, this distinction is often discussed in the form of traits and states. Personal traits constitute those "relatively stable interindividual differences in the degree/extent/level of coherent behaviours, thoughts, and feelings" whereas states refer to "the degree/extent/level of coherent behaviours, thoughts, and feelings at a particular time" (Baumert et al., 2017, p. 528). States are thus actual "occurrences" (p. 172) of affects, behaviours, cognitions, or motivation whereas traits are "dispositions" (p. 172) towards certain states (Fridhandler, 1986). Hence, whilst states are short-lived experiences that may fluctuate across different situations (e.g., Schmitz, 2006), traits are relatively more enduring personality aspects that can remain stable for as long as entire life spans (Baumert et al., 2017; Fridhandler, 1986). The reason that states are rather short-lived phenomena is that they form realisations of a person's dispositions within a concrete situation, but they do so reflecting those psychologically salient contextual factors that make up a specific situation (Fleeson, 2001; Fridhandler, 1986). Hence, any change within the situational context most likely will lead to new realisations of a person's psychological traits and as a result the person may additionally experience a change in their psychological states. With regards to traits, McCrae and Costa (e.g., 1996, 1999) further distinguish between basic tendencies and context-specific characteristic adaptations. The former constitute biologically determined dispositions that are "endogenous" (McCrae & Costa, 1999, p. 144) in the sense that they are not subject to environmental influences of any form. The latter are defined as “acquired skills, habits, beliefs, roles, and relationships” (McCrae et al., 2000, p. 174) that evolve in life domains (McAdams & Pals, 2006) through the interaction between traits and environment (McCrae & Costa, 1996). Within education, characteristic adaptions manifest themselves especially in the form of academic beliefs. Academic beliefs are defined as “the implicitly or explicitly held subjective conceptions” (Eynde et al., 2002, p. 28) individuals consider true about education and about themselves in relation to education. Aacdemic beliefs are subjectuve in nature that emerge and increasingly solidify over the course of those years spent in school. Conceptually speaking, they are located between traits and states at the level of school subjects (Boekaerts, 1996; see also Nicholls, 1979; Shavelson et al., 1976; Shavelson & Bolus, 1982). School subject-specific beliefs constitute a “student’s tendency to react favourably or unfavourably to learning in that domain” (Boekaerts, 1996, p. 395). Within the literature, different motivational constructs have been conceptualised hierarchically using these levels. For example, Shavelson et al. (1976) presented a multilevel conceptualisation of a person’s self-concepts. At the top of the hierarchy, they locate a general but stable self-concept (i.e., trait) whereas they locate situation-specific experiences of competence (i.e., states) at the bottom. In between, they place school subject-specific self-beliefs, which they call “subareas of self-concept” (p. 413). A similar conceptualisation has been proposed for enjoyment (Götz et al., 2006), and self-determination (Vallerand, 1997). Education-related subjective conceptions may not only operate at the school subject-specific level. As a matter of fact, researchers have studied achievement motivation also at the lesson-specific level (Ahmed, Minnaert, et al., 2010; Böheim et al., 2020; Tsai et al., 2008; Vasalampi et al., 2020). Figure 1 on the next page gives an example. It depicts a multilevel representation of expectancy, value, or cost (self-) evaluations that spans both the school-subject and the lesson-specific level. It highlights two aspects. First, the multidimensional structure of achievement motivation may not necessarily vary across levels of specificity. At both the lesson-specific and school subject-specific level, achievement motivation comprises of expectancy, value, or cost (self-) evaluations. Second, both the lesson-specific and school subject-specific level vary in terms of their "generality" (Skinner, 1996, p. 555). (Self-) evaluations that are lesson-specific may just apply to a specific lesson (see lessons 1 to 3 in figure 1) and thus may vary across lessons. By contrast, school subject-specific (self-) evaluations may hold not only at the time of a given lesson but may generally hold across different lessons. In fact, whereas states have been shown to vary across days, beliefs have been found to be rather stable within (e.g. Marsh et al., 2005; study 1) and across school years (Lauermann et al., 2017; Marsh & Martin, 2011; Pinxten et al., 2014).

Figure 1: Multileveled achievement motivation spanning the lesson and school level



The intrapersonal variability associated with states has been documented within the motivational literature. A two-level analysis[4] of state data ("variance component models")

[4] As section 2.1.2 highlights, higher-level models to analyse constructs are feasible. A two-level model is just one analysis frame used to better illustrate the point made.

allows researchers to attribute the relative variance within a construct to the state and trait level (i.e., intraclass correlations (ICCs); Raudenbush & Bryk, 2002). The ICC expresses the relative variance of a latent variable across levels and thus lies between 0 to 1 (Raudenbush & Bryk, 2002). In case ICCs substantially differ from 0, multilevel analysis is required in order to control for the clusteredness of estimated standard errors (Goldstein, 2011; Raudenbush & Bryk, 2002). Any ICC approaching 0 indicates dependence of motivational experiences especially on situation-specific factors rather than person-specific characteristics (Kreft & Leeuw, 1998). Studies partitioning variances into trait and state components found motivational states to vary substantially across situations. ICCs reported in the literature are between .27 and .44 in the of expectancy-like constructs (e.g., Ahmed, Werf, et al., 2010; Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021), between .19 and .36 in the case of value beliefs (e.g., Dietrich et al., 2019; Salmela-Aro et al., 2021), and between .28 and .55 in case of cost perceptions (e.g., Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021). This implies that between 45 and 72 per cent of the variance is at the state level. Hence, any representation of achievement motivation must take into account the situational dependency of its actual experiences.

2.1.3. Empirical operationalisation of achievement motivation

Motivational experiences are generally measured through self-reports (Appleton et al., 2006). After all, beliefs constitute person-internal experiences that cannot be directly observed. To construct motivational measures, researchers have to take into account several design choices (Bolger & Laurenceau, 2013). One design feature concerns whether a self-report item refers to a specific task/event or to broader contexts such as a school subject

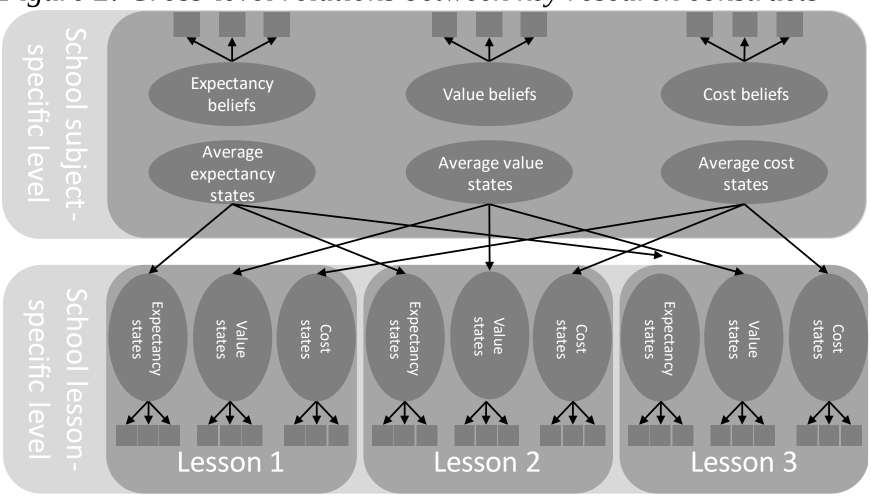
(Bong & Skaalvik, 2003). To measure task-related self-efficacy (e.g., Bandura, 1986) for example, Marsh et al. (2019) handed out mathematics problems and then asked respondents "How confident are you that you can solve *this math problem*?" (p. 340; emphasis added). To give task-related self-appraisals, respondents need to ask themselves whether or not they believe they can solve the given problem (i.e., the external referent). Especially within quantitative school subjects, this is a yes/no question. Either one obtains the solution or not. In contrast to task-related instruments, measures that are especially episode- or school subject-specific tend to exist without a concrete referent. Examples from the lesson-specific level include "How *well* did you understand *what was taught*?" (Vasalampi et al., 2016, p. 48; emphasis added). Since no explicit referents are provided, these measures are "evaluative" (Marsh, Pekrun, et al., 2019, p. 335) in nature. After all, whether or not something was understood *well* may be relative. To measure motivational variables at the school subject-specific level, researchers ask respondents about what they think their typical motivational experiences are within a given school subject such as mathematics or biology. Examples of measures used to gauge motivational beliefs include items such as "How good at [SUBJECT] are you?" to measure expectancy (i.e., self-concept) beliefs and "In general, how hard is [SUBJECT] for you?" (Eccles & Wigfield, 1995, p. 224) to measure cost beliefs. Whether or not self-report items are task-/event-specific and thus contain an explicit referent may have important implications, which will be discussed further in section 2.1.3. Items that are not situation-specific have another characteristic that may distinguish them from task-specific items. Especially school subject-specific items require aggregation of individual experiences into one overall rating. In order to respond to them, respondents are assumed to aggregate all their previous experiences within the episode or subject concerned into one overall rating (i.e., time-inclusive reports; Robinson & Clore, 2002). This may be cognitively difficult if not impossible. It would require recalling all those events that should be taken into account when determining one school subject-specific rating (Blair & Burton, 1987; Fleeson & Gallagher, 2009; Patrick & Zuckerman, 1977). To circumvent those cognitive challenges, respondents of school subject-specific self-reports are assumed to revert back to decontextualised generalisations about themselves or the subject that they have accumulated over time (i.e., semantic memory; Robinson & Clore, 2002). However, because they reflect people's generalisation, it has been argued that school subject-specific self-reports are best understood as subjective beliefs (self-reported motivational beliefs) about one's typical experience within a given subject rather than an accurate indication of those actual experiences one typically has in that subject (e.g., Bieg et al., 2014; Götz et al., 2013; Roos et al., 2015). For example, school subject-specific self-reports seem to suggest that students exhibit stronger emotions within academic domains such as mathematics than they actually do when being asked right in the moment by the means of state self-reports (e.g., Bieg et al., 2014). State self-reports are generally about one specific event (i.e., the state) and are normally also administered in close proximity to the event of interest. Thus, when responding to state self-reports respondents can access their cognitive-affective experiences straight away ("experiential knowledge"; Robinson & Clore, 2002, p. 934) and are also not required to aggregate a number of cognitive-affective experiences into one overall rating.

Given the different methodological advantages of state self-reports, it has been proposed (Beal, 2015; Fleeson, 2009; Fleeson & Gallagher, 2009; Mischel & Shoda, 1995; Shoda et al., 1994) to rely on them when inferring stable person characteristics such as traits. The underlying idea is not to have the respondent aggregate their different experiences into one overall rating (i.e., the trait) but to aggregate the respondent's different state self-reports into one overall score. Accordingly, a trait is modelled as the person-specific average that aggregates a series of situation-specific self-reports into one summary statistic.[5] In terms of the self-reports including a large enough collection of different situational contexts (Fleeson & Gallagher, 2009), person averages should then be a "robust" (Beal, 2015, p. 398) measure of people's characteristics. To ensure a large enough collection of different situational contexts, states are measured repeatedly within relatively short timeframes (i.e., intraindividual/intensive longitudinal design; e.g., Bolger & Laurenceau, 2013). Person characteristics such as traits or motivational beliefs can then be inferred based on repeated self-reports (i.e, intensive longitudinal data; Bolger & Laurenceau, 2013) through the application of latent state/ trait models (e.g., Castro-Alvarez et al., 2021; Geiser, 2013, 2020). Within single-trait multiple state (STMS) models, the guiding assumption is that a person's measurement scores observed longitudinally can be decomposed into a person-specific average (i.e., the trait) and time point-specific deviations (i.e., states) from that time-invariant average (Castro-Alvarez et al., 2021; Geiser, 2020; Geiser et al., 2013). Following this logic, school subject-specific motivational can be gauged not only through subject-specific self-reports but also through averages of state self-reports taken repeatedly over time within the subject being studied. In that scenario, as figure 2 below highlights, self-reported motivational beliefs and average motivational states match each other in terms of their level of specificity. Within personality research, it has been discussed to what extent average states and direct self-reports constitute methodologically different but conceptually identical representations of stable person characteristics (Augustine & Larsen, 2012; Breil et al., 2022; Finnigan & Vazire, 2018; Fleeson & Gallagher, 2009; Rauthmann et al., 2020). To address this question, researchers have been examining their interrelations (e.g., Augustine & Larsen, 2012; Fleeson & Gallagher, 2009) as well as the congruence within their relations with other

[5] Also see Fleeson and Gallagher (2009) for the usage of other measures of central tendency to model latent traits.

external variables (Breil et al., 2022; Finnigan & Vazire, 2018; Rauthmann et al., 2020). The latter then implies external validity. The former constitutes convergent validity, which is about the degree to which different operationalisations of a construct, or overlapping constructs, are interrelated (Bryant, 2002; e.g., Crocker & Algina, 2006; Irwing et al., 2018). Convergent validity has often been concluded when (factor) correlations between two operationalisations of interest are extremely high (Raykov & Marcoulides, 2011). Within personality research, the growing consensus has been that both global and average measures are related but distinct constructs (e.g., Augustine & Larsen, 2012; Fleeson & Gallagher, 2009). First evidence from the educational domain suggests a similar verdict. For example, Ahmed et al. (2010) and Malmberg et al. (2013) find only moderate correlations (~.43) between self-reported motivational beliefs and average motivational states. Others (e.g., Bong, 1998; Marsh, Pekrun, et al., 2019) find self-reported motivational beliefs and average motivational states to be differentially predicted by prior academic achievement. However, research comparing self-reported motivational beliefs and average states appears to be sporadic within the educational domain. Furthermore, this research is limited exclusively to the Western world excluding other regions such as Africa.

Figure 2: Cross-level relations between key research constructs



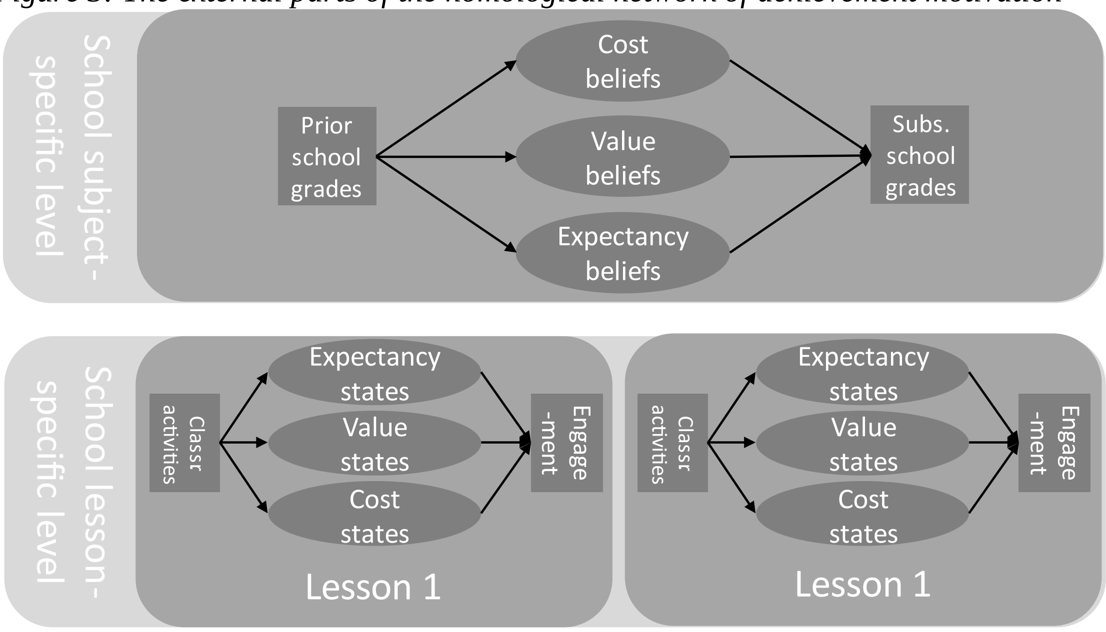
2.2. The external nomological net around achievement motivation

In section 2.2, the literature on key components of the external nomological network around achievement motivation is discussed. The external nomological network consists of those relations between the construct of interest and other constructs that are assumed to be mutually distinct (Byrne, 1984). In the case of achievement motivation, this concerns relations with external constructs presumed to be predicted by it (i.e., criterion variables) and relations with external constructs presumed to predict achievement motivation (i.e., predictor variables). As discussed in section 2.1.2, achievement motivation is a multilevel concept. Hence, external constructs may relate to achievement motivation at the different levels of specificity. In line with the "specificity matching principle" (Swann et al., 2007, p. 87), it is therefore important to relate criterion and predictor variables to achievement motivation at the "right" level of specificity. According to the specificity matching principle, predictor variables explain only those outcomes that are equally specific (Swann et al., 2007). For example, a lesson-specific variable may predict an outcome that is also lesson-specific, but it may not predict a state that is measured at any given moment in time during a lesson. To provide the reader with orientation for sub-section 2.2.1 through 2.2.6, the figure 1 has been extended to now include those predictor and criterion variables that are studied as part of achievement motivation's external nomological network in Rwanda (see figure 3 on the next page).

Sub-sections 2.2.1 and 2.2.2 introduce academic achievement and engagement as outcomes hypothesised to be predicted by achievement motivation. Together, both sub-sections can also be taken as an indication of the practical relevance of achievement motivation. As far as achievement motivation predicts relevant educational outcomes, its study across different contexts may also be justified from a policy perspective. Sub-sections

2.2.3 and 2.2.5 introduce academic achievement, classroom activities as well as climate as predictors of achievement motivation. Within the literature, reciprocal relations between academic achievement and achievement motivation have been highlighted (e.g., Niepel et al., 2014). This means they influence each other. The initial plan was to study the reciprocal relations between academic achievement and achievement motivation. However, soon after the completion of the motivational data collection in February 2020, the Covid pandemic broke out in Rwanda. Consequently, all schools were closed down and students did not take any exams. It was therefore not possible to study the effects of achievement motivation on academic achievement. To obtain at least some insights into their relations, the study of effects of academic achievement on achievement motivation was therefore particularly valuable within the context of this DPhil project. Section 2.2.6 introduces key covariates, which the study of achievement motivation within education contexts should consider.

Figure 3: The external parts of the nomological network of achievement motivation



2.2.1. Achievement motivation as a predictor of academic achievement

The educational relevance of achievement motivation may by and large depend on the following two aspects: first, to what degree it predicts important educational outcomes such as academic achievement. Second, to what degree achievement motivation is malleable in the sense that students' motivational experiences can be altered through targeted interventions. The motivational constructs outlined within expectancy-value theory have been related empirically to important educational outcomes, including academic achievement. Studies on the relations between motivational variables and academic achievement are particularly informative if they are able to isolate additional effects of motivational variables over and beyond the influence of control variables such as intelligence or prior achievement. Studies on the incremental effects generally find expectancy beliefs to be an important predictor of achievement (Hailikari et al., 2008; e.g., Kriegbaum et al., 2018; Lavrijsen et al., 2021; Steinmayr & Spinath, 2009). Using a longitudinal design with a sample of around 3000 Belgian secondary students, Lavrijsen et al. (2021) estimated the incremental effects of a number of different motivational constructs including school-level expectancies (i.e., general self-concept), all measured at the beginning of a school term. Learning achievement was measured through both school marks (point average) as well as standardised tests in mathematics, both taken at the end of the school term. Controls included intelligence as well as prior academic achievement. To determine the relative importance of predictors (i.e., relative weight analysis), Lavrijsen et al. estimated their contribution to the overall coefficient of determination attained within regression models (the sum of each contribution was then equal to the model's overall R^2). The full model with average school marks as the outcome explained around 86 per cent of the variance. Prior achievement was by far the biggest predictor with around 58 per cent whereas all motivation variables jointly explained only around 16 per cent. However, self-concept was the second biggest predictor with around 11 per cent and thus even larger than intelligence (around nine per cent). This result was not fully replicated in models using standardised test scores as dependent variable (R^2 = 52.36 per cent; prior achievement was not controlled for). Here, intelligence was by far the biggest predictor (around 35 per cent) but then self-concept was again the second biggest factor (around 7 per cent).

In terms of value perceptions, results are mixed. Steinmayr und Spinath (2009) examined relative weights associated with value beliefs. Using a sample of 342 German upper-secondary students, end-of-year school marks were regressed on motivational variables, including expectancy and value beliefs, while controlling for both intelligence and school marks from the previous year. Within the mathematics domain, the model predicted achievement well ($R^2 = 0.58$). Again, prior achievement accounted for the largest portion of the explained variance (44.6 per cent). Besides, however, Bulla et al. found not only expectancies (22.4 per cent) but also value beliefs (18.6 per cent) to be important predictors. By contrast, intelligence was associated with only 5.1 per cent of the explained variance. Interestingly, within the subject of German (i.e., language domain), results were slightly different. Both expectancies (12.5 per cent) and value (8.1 per cent) accounted for smaller portions of the explained variance ($R^2 = 0.22$) while prior achievement (50.7 per cent) and intelligence (11.9 per cent) turned out to be more important.

Compared to Steinmayr und Spinath, Kriegbaum and colleagues (2018) found weaker effects associated with value beliefs over and beyond intelligence. To determine the incremental effects of motivation on school achievement (operationalised through school marks or standardised test results) over and beyond intelligence, they performed meta-analytic regressions that aggregated results from 74 individual studies identified. Motivation was included as a single predictor but the nature of motivational variables within single studies was accounted for through different dummy indicators (i.e., moderators) included in the model. Using this method, Kriegbaum et al. could compare the relative importance of expectancy and value belief. Expectancy beliefs turned out to be the strongest predictor amongst all motivational constructs, accounting for around 30 per cent of the explained variance ($R^2 = 0.3$) while intelligence contributed around 47 per cent. Value effects were modest in comparison. It accounted for only about nine per cent of the total variance explained by the model ($R^2 = 0.23$) while intelligence accounted for around 78 per cent. A major limitation of studies on the incremental effects was the unidimensional operationalisation of value even though it constitutes a multidimensional construct. In fact, zero-order correlations with academic achievement suggest that intrinsic value is a significant predictor (Ozel et al., 2013). Comparable studies were not performed involving cost constructs. However, investigations of its predictive power within confirmatory factor analyses show that that costs are related to school performance (Flake et al., 2015; Kosovich et al., 2015; Perez et al., 2019).

In comparison to the importance of competence and intelligence for academic achievement, the effects of motivation appear small. This may partly be due to the way incremental-effects models are conceptualised. Whereas intelligence is rather a stable construct (Protzko, 2015), academic achievement, especially when operationalised through school grades, interacts with motivation across time. Motivation influences academic achievement but is also influenced by academic achievement (Huang, 2011).[6] Hence, the relative weights of motivational effects in models that control for prior academic achievement are underestimated if these models do not take into account that prior achievement levels in turn have been influenced by previous motivational levels (Lavrijsen et al., 2021; Steinmayr & Spinath, 2009). In fact, it has been argued that motivation may unfold over time through "virtuous and vicious feedback cycles" (Pitzer & Skinner, 2017, p. 24). Accordingly, students

[6] Respective relations are best studied using reciprocal effects models (e.g., Marsh & Yeung, 1997).

that are motivated and thus experience academic success may become even more motivated. By contrast, students that have been struggling from a motivational point of view may become increasingly disengaged, which further reduces their achievement motivation for school. These processes, however, are not set in stone. Unlike intelligence, motivation has been found to be malleable. Targeted interventions that aim to alter students' expectancy and value perceptions have indeed been found to positively influence students' motivational levels (e.g., Hulleman, Godes, et al., 2010; Lazowski & Hulleman, 2016; Yeager & Walton, 2011). Within the value component of the expectancy-value conceptualisation of achievement motivation, utility perceptions appear to be more susceptible to the effects of external interventions that intrinsic and attainment value. Evaluation studies have shown that utility value can be shifted through targeted interventions with positive knock-on effects on achievement (e.g., Harackiewicz et al., 2016; Hulleman, Godes, et al., 2010). However, achievement motivation is not expected to exert direct effects on academic achievement. Its effects are rather expected to be mediated by behavioural engagement. This outcome variable is introduced next.

2.2.2. Achievement motivation as a predictor of engagement

Achievement motivation refers to the psychological processes that ultimately shape student actions exhibited within education-relevant learning and performance contexts (Wigfield et al., 2006, 2015). Hence, student engagement (i.e., student actions) has been built into the very fabric of the theoretical conceptualisation of achievement motivation. In fact, achievement motivation has been called the "antecedent of engagement" (Hospel et al., 2016, p. 38) while student engagement is labelled "the outward manifestation of motivation" (Skinner et al., 2009, p. 224). Student engagement is defined as "student's active participation in academic and cocurricular or school-related activities, and commitment to educational goals and learning” (Christenson et al., 2012, p. 816). It has been studied at different levels. At the general level, engagement refers to the involvement of students in those institutions that foster their socialisation into society. At this level, school is just one yet an important prosocial institution (Skinner & Pitzer, 2012). At the school level, engagement refers to students’ participation in academic and non-academic activities (Li & Lerner, 2011) that are beneficial for the development of peer relations. Engagement at the classroom level refers to the extent to which students are focussed on lesson activities (Skinner & Pitzer, 2012) and has been defined as “participation in academic work” (Skinner, 2016, p. 146). Classroom-based student engagement is particularly crucial for students’ academic success since it is predominantly through lesson activities that students build up those knowledge and skills required for their school careers. In fact, it has been considered “the direct (and only) pathway” (Skinner & Pitzer, 2012, p. 23) to learning acquisition that is available to students. This in turn explains why engagement is considered to mediate the effects of achievement motivation on academic achievement.

Classroom-based engagement has been conceptualised as a three-dimensional construct (Fredricks et al., 2004). One dimension refers to the affective reactions students experience in classroom contexts (i.e., “emotional engagement”; Fredricks et al., 2004; Skinner et al., 2008). Another dimension refers to the extent to which students are nondisruptive and pay attention to classroom activities (i.e., “behavioural engagement”; Fredricks et al., 2004; Olivier et al., 2020; Skinner et al., 2008). A third dimension refers to whether or not students guide their classroom efforts through self-regulated learning strategies (i.e., “cognitive engagement”; Fredricks et al., 2004; Zimmerman, 1990). Lei et al. (2018) carried out a meta-analysis in which they pooled 69 studies on the nexus between each of the three components of student engagement and achievement outcomes. With an average weighted effect size of around $r = .350$ (95% $CI = .258, .279$), behavioural engagement appears to predict academic progress to a larger extent than cognitive ($r = .245$; 95% $CI = .235, .256$) and especially emotional engagement ($r = .216$; 95% $CI = .208, .224$). Not only does behavioural engagement appear to be a stronger predictor of academic achievement than cognitive and emotional engagement, but it may also be the necessary requirement for students to get involved in academic activity cognitively and emotionally in the first place. For example, only if a student starts working on a task, (s)he may start experiencing joy about it.

Behavioural engagement has been discussed in terms of on- and off-task behaviour. It is centred around the hypothesis that the longer one pays attention to an activity the better the learning outcomes (e.g., Carrol, 1963; Wiley & Harnischfeger, 1974). Being off-task has been defined as "any behaviour that does not involve the learning task or material, or where learning from the material is not the primary goal" (Rodrigo et al., 2013, p. 2) whereas on-task behaviour is understood as a "state in which attention is directed more or less exclusively to one target or task" (Ruff & Rothbart, 2001, p. 110). Based on a review of the corresponding literature, Hospel et al. (2016) developed items reflecting the different aspects associated with behavioural engagement. Their exploratory and confirmatory factor analysis revealed a five-dimensional structure consisting of being in class ("absenteeism"), being ill-behaved ("disruptive behaviours"), being off-task ("withdrawal"), being on-task ("following instructions") and being verbally engaged ("participation"). These five dimensions can also be interpreted as different levels of behavioural engagement. The most visible indication of disengagement is absenteeism, which implies students not being in class in the first place (Reid, 2005; Teasley, 2004). Off-task behaviour within classrooms appears either in the form of disruptive misconduct (e.g., talking loudly with a neighbour about something unrelated to school) or in the form of passive inattention (e.g., dozing off in one's seat; Finn et al., 1995).

By contrast, on-task behaviour in class is exhibited either in the shape of passive but focussed involvement with a task or through focussed and active participation in learning activities. Whatever the way of operationalising it, time on-task has been found to be a predictor of learning outcomes (Connor et al., 2009; Fisher et al., 2014; Gobel, 2008; Godwin et al., 2021; Guillaume & Khachikian, 2011; Kupiainen et al., 2014; Moffett & Morrison, 2020). Empirical evidence also supports the hypothesis that motivation predicts behavioural engagement (Dietrich et al., 2017; Garn et al., 2017; Luo et al., 2016; Pekrun et al., 2017; Putwain et al., 2018). For example, Patall and colleagues (2016) analysed relations between lesson-specific intrinsic value ("situational interest") and three measures of behavioural engagement (e.g. "I paid attention today in class") amongst secondary students. For six weeks, respondents reported on each measure after each weekly science lesson. Given the clustered nature of the data in the sense that repeated measures (level 1) were nested within individuals (level 2), the researchers ran hierarchical regression models (see below for a more detailed discussion of nested data structures). All estimated coefficients were statistically significant (standardised $\beta' s$ between $.17$ and $.23, all\ p's <$ .001). Hence, one standard-deviation increase in enjoyment was associated with an increase of around .17 to .23 standard deviations in engagement (standardised βs between $.17$ and $.23, all\ p's <$ .001). Patall et al. (2016) also controlled for prior engagement levels ("lagged outcome") in their models, which further increases confidence in the educational relevance of students' lesson-specific enjoyment. Using a similar study design, Dietrich et al. (2017) found that the omnibus variable value that combined self-report items on intrinsic, utility, and attainment value predicted situation-specific behavioural engagement ("effort"). They also found that situation-specific expectancies were significantly related to effort. Conversely, costs appear to be a substantial but weak determinant of effort. Hence, achievement motivation is relevant from an educational point of view because it predicts academic achievement. However, its effects on academic achievement are mediated especially by behavioural engagement.

2.2.3. Academic achievement as a predictor of achievement motivation

As already alluded to, to effectively estimate the effects of achievement motivation on academic achievement one needs to control for prior academic achievement. This highlights that achievement motivation and academic achievement are reciprocally related. Different models have been developed to explain how prior academic achievement shapes academic achievement. One of the most eminent models in this respect is the generalised internal/external (GI/E) model, which has sparked a large number of studies (Möller et al., 2009, 2020) across the world (e.g., Marsh & Hau, 2004). The GI/E model is an extension of the classic I/E model (e.g., Marsh, 1986), which explains the levels of self-concept exhibited within the mathematics and language domain. Both models thus operate at the school subject-specific, or domain-specific for that matter, level.[7] It hypothesises that self-concept in each domain is based on self-evaluations carried out using the following two comparison points. First, students evaluate their own academic achievement in any of the two domains by contrasting it with achievements attained by significant others (e.g., classmates) in that domain (i.e., external references). Within external (i.e., social) comparison processes, higher academic achievement should lead to higher levels of competence beliefs. Second, students evaluate their achievement in a given domain by contrasting it with achievement they have

[7] A school domain (e.g., language domain) is a content area that may comprise of not only one but several different school subjects (e.g., English and Kinyarwanda). Domain specificity may imply that a given motivational construct holds across all subjects within that domain. Domain specificity is expressed especially through high correlations of this construct across the different subjects within the domain being studied. School subject-specific thus implies that a given motivational construct holds across the subject being studied but not necessarily across all subjects within the corresponding domain. As will be discussed later, motivational constructs often do not exhibit domain specificity within languages.

attained in the other domain (i.e., internal reference). Internal (dimensional) comparisons are a "double-edged sword" (Möller & Marsh, 2013, p. 546) because they lead to higher self-concept within the domain that is associated with relatively higher performance while leading to lower self-concept in the domain that is associated with relatively lower academic performance. Hence, the classic I/E model explains the paradoxical near-zero correlations between mathematics and language self-concepts through the interaction of the counteracting processes of external and internal comparisons (Möller et al., 2009, 2020). The GI/E model of motivational beliefs applies the I/E model logic to other motivational beliefs than only self-concept and to other school subjects than only mathematics and languages (Möller et al., 2016). Model extensions to other school subjects were mainly guided by dimensional-comparison theory (Möller & Marsh, 2013). The theory implies that school subjects can be arranged along a continuum based on their perceived level of similarity. Mathematics and language form the opposing ends of that continuum. All other school subjects can then be positioned within this verbal-to-mathematics continuum based on their perceived similarity/difference with both ends (Marsh et al., 1988). Insofar as the two subjects are different (i.e., "far" comparisons), students are assumed to explain performance between both domains through different sets of abilities. Insofar as two subjects are seen to be similar (i.e., "near" comparisons), students are assumed to explain performance in any of the two subjects through the same set of abilities. Within far comparisons, the effects of prior achievement in one subject on motivational beliefs in the other subject are expected to be negative (i.e., contrast effects). Within near comparisons, the cross-domain effects are expected to be positive (i.e., assimilation effects; Marsh, Lüdtke, et al., 2015; Möller & Marsh, 2013). For example, Jansen et al. (2015) found assimilation effects between chemistry and mathematics but contrast effects between chemistry and languages. According to dimensional comparison theory (Möller & Marsh, 2013), one would expect assimilation effects between subjects within the language domain. However, only some empirical investigations found assimilation effects (e.g., Marsh et al., 2014; Möller et al., 2006), while a large number of studies observed contrast effects (e.g., Arens et al., 2020; Marsh et al., 2001, 2014; Marsh, Lüdtke, et al., 2015; Niepel et al., 2014; Stocker et al., 2021; van der Westhuizen et al., 2022), or no effects at all (e.g., Gaspard et al., 2018; Xu et al., 2013). One explanation of these contradicting results could be rooted in the salience of languages being compared. The more the two languages compared are integral parts of students' lives, as may be common in multilingual contexts, the more students are aware of their differences, which in turn may lead to contrast effects between them (van der Westhuizen et al., 2022).

Extensions of the classic I/E model to other motivational beliefs concerned utility value and cost beliefs (Gaspard et al., 2018; Guo et al., 2017). Comparing different languages and quantitative subjects, Gaspard et al. (2018) found that higher school grades were associated with more favourable cost perceptions within domains. Between quantitative and language subjects, cost beliefs were subject to contrast effects. Between quantitative subjects, cost beliefs were subject to assimilation effects. As for utility value, they only detected within-domain effects, which were significantly weaker, especially in comparison to those observed for expectancy and cost beliefs (also see Guo et al., 2017). Compared to other value facets (Götz et al., 2008; Nagy et al., 2006; van der Westhuizen et al., 2022), however, utility perceptions have been subjected to GI/E model research to a much lesser extent. More research is therefore needed to examine its relatedness to prior achievement not only in the Western world but also in Sub-Saharan Africa.

A review of the GI/E model literature reveals two important aspects. First, Chiu (2008) has been the only study to test the model within the Africa context, an aspect that will be discussed further in section 4.3. Second, GI/E model studies have mainly relied on measuring motivational beliefs through school subject-specific self-reports. As highlighted by figure 2 on page 39 and further elaborated in section 2.1.3, average motivational states may constitute a different, potentially more adequate way to represent typical motivational experiences within a school subject. With only a few exceptions (Bong, 1998; Marsh et al., 1991; Marsh, Pekrun, et al., 2019; Skaalvik & Rankin, 1995), the extent to which prior achievement also predicts motivational beliefs operationalised through methodologies comparable to average motivational states appear to have hardly been studied. A key characteristic of the few studies that included constructs comparable to average motivational states focussed on self-efficacy, which they measured through task-related items that provide a clear referent. To provide a task-related self-efficacy rating, respondents just need to ask themselves whether or not they believe they can solve the problem specified (i.e., the external referent). Item responses of respondents may be influenced by their prior experiences with similar tasks (Bandura, 1986). However, they may not have the need to engage especially with dimensional comparisons to determine whether or not they feel capable of solving the given problem. Task experiences from other school subjects such as language may just not help to provide an adequate item response. It has therefore been argued that internal frame-of-reference effects in particular may play a lesser role in explaining motivational traits that are gauged through test-related items (Marsh, Pekrun, et al., 2019). GI/E model studies on test-related self-efficacy (Bong, 1998; Marsh et al., 1991; Marsh, Pekrun, et al., 2019; Skaalvik & Rankin, 1995) supported this hypothesis, which made Möller et al. (2009) conclude that “the I/E model is valid for self-concept but not for self-efficacy beliefs” (p. 1158). However, motivational states have also been measured through items that do not provide an explicit referent and thus are evaluative in nature as exemplified by items such as “How *well* did you understand what was taught?” (Vasalampi et al., 2016, p. 48; emphasis added). To come up with answers, students may indeed revert to social and even dimensional comparisons. After all, whether or not something was understood *well* may be relative. If this holds true, then school subject-specific motivational beliefs, conceptualised as average states gauged through episode-specific items, may indeed be explained through the GI/E model logic as well. Hence, whereas convergent validity between school subject-specific motivational beliefs and average motivational states can be investigated through their correlations, the application of the GI/E model to both school subject-specific motivational beliefs and average motivational states can be seen as a test of their external validity.

2.2.4. Classroom activities as predictors of achievement motivation

Achievement motivation also concerns learning in action. Within school education, the classroom is a key location where learning takes place. To the extent that motivational states are influenced by situational factors, what is happening within classrooms should thus have a direct bearing on motivation. Within educational research, different frameworks have been proposed to capture classroom-level factors. For example, within self-determination theory, classroom-level factors have been analysed in terms of the how clearly expectations are communicated (structure), the level of choice provided (autonomy support), and the level of warmth expressed (involvement) by the teacher (e.g., Jang et al., 2010). However, these factors may be best understood as classroom-climatic factors. They may vary between classes but not necessarily between learning situations within classes (e.g., Marsh et al., 2012; Morin et al., 2014). For example, once students have internalised a teacher to be involved then this perception may be rather stable. They may have to behave in an uninvolved way repeatedly until this new impression starts to sink in with students. On the other hand, classroom activities may be classroom-level factors that may truly vary across situations. Classroom activities constitute situation-specific “opportunities to learn” (Jansen, Scherer, et al., 2015, p. 15) that students engage with during a lesson. They are situation-specific because oftentimes students do not engage with just one classroom activity over the course of a lesson. A lesson may start with some classroom discussions, then continue with some note taking and individual seatwork, and may finally end with some group discussions of the individual seatwork. Within the international-development discourse on education, the classroom activities have been discussed using the lens of student-centred versus teacher-centred classroom practices (e.g., Bremner et al., 2022; Schweisfurth, 2013). For example, the launch of Education for All (EFA) in 1990 embodied an agreement by international-donor agencies that the most effective pedagogies would be "learner-centred" (Haddad et al., 1990, p. 68). Ever since, educational reforms in Africa (e.g., Frost & Little, 2014; O'Sullivan, 2004; Sakata, 2022; van de Kuilen et al., 2019) and elsewhere (e.g., Bremner et al., 2022) have stressed student-centred activities. Within student-centred approaches, students are the main actors within learning processes who take responsibility for their own learning progress. Accordingly, teachers assume a predominantly faciliatory role within classrooms in the sense that they help to adjust learning processes to students' prior knowledge and current needs (Bremner, 2021). By contrast, teacher-centred approaches are organised around the teacher who controls the pace and content of the learning process in the form of whole-class instructions often derived from textbooks. Within these whole-class processes, students are mainly passive recipients of knowledge imparted by the teacher (Smit et al., 2014). Arguments put forward in support of student-centred approaches implied that they help students to learn more effectively, to develop critical thinking as well as creativity, and to practise principles that have been otherwise associated with democratic citizenship (Schweisfurth, 2013). A first review indeed suggests generally positive cognitive as well as non-cognitive outcomes associated with student-centred pedagogies within low- and middle-income countries (Bremner et al., 2022). However, the authors also point out that the existing evidence is "provisional" at best and that more research on the effects of student-centred classroom practices within low-resource countries is warranted. Current research on student-centred pedagogies may also be hampered by the fact that they constitute complex concepts (Bremner, 2021; Schweisfurth, 2013). Studying their various aspects in isolation may help to identify what elements of student-centred pedagogies are most relevant to student outcomes. One of most common principles associated with student-centred pedagogies has been called “interaction” (Bremner, 2021, p. 166), “dialogue” (Schweisfurth, 2013, p. 146), or “active teaching” (Ngware et al., 2012), and refers to the provision of opportunities to students to verbally interact with each other or with the teacher during lessons (Bremner, 2021). As the next section highlights, they have also been found to predict motivational variables.

Verbal interactions within classrooms appear also in the form of whole-class discussions and group work. The former can be loosely defined as “all verbal exchanges where one individual addresses another individual or individuals and at least one addressed individual replies” (Howe & Abedin, 2013, p. 325). It is implied that through classroom discussions students feel they have a say in learning activities and thus can contribute to their own learning process. Discussions can thus be a strategy to improve students’ autonomy (Deci & Ryan, 2000), a state which has been associated with feelings of competence and interest. The former resonates with Jansen et al. (2015), who found classroom discussions to positively predict students’ competence experiences within science lessons. It has been argued that especially situational interest positively influences students’ attentional and retrieval processes (Hidi, 1990), which in turn should reduce cost perceptions around learning activities. Small-group work implies students being divided into groups where they work together to complete a given task through verbal interactions (e.g., Murphy et al., 2009). Compared to whole-class discussions, perceived autonomy may thus even be larger within small-group settings, with positive effects on students’ interest levels (X. Wu et al., 2013) and thus cost perceptions. In small-group settings, individuals that work together can pool their mental resources and abilities, which should lead to higher confidence levels of being able to successfully complete the task (i.e., group efficacy; Bandura, 2000) compared to settings where students work individually. Furthermore, especially when performing challenging tasks, working in groups requires less mental effort from its members in order to attain given learning outcomes (i.e., "efficiency"; F. Kirschner et al., 2009, p. 615). Hence, incorporating group work into lessons may also lead to more favourable competence and cost (self-) evaluations amongst students. Both classroom discussions and group work may also have positive effects on utility perceptions around lesson content. As already pointed out, utility perceptions are especially shaped through significant others such as teachers and peers (Harackiewicz et al., 2012b; Wigfield et al., 2020). Classroom interactions may thus provide students with the opportunity to enquire about the instrumental value of lesson content. It might have been for this reason that Urdan and Turner (2005) make the case for "active participation" (p. 303) in classrooms as a way to promote the perceived task value of learning.

Despite a policy impetus towards student-centred learning, teacher-centred practices have tended to persist in many low- and middle-income countries (Nag, 2014; Schweisfurth, 2013). Using classroom observations, Sakata (2022) found teacher-centred activities to account for 81 per cent of the lesson time at Tanzanian primary schools whereas student-centred activities accounted only for about 14 per cent. Frost and Little (2014) found teacher-oriented activities accounted for around 75 per cent and student-centred activities for another 15 per cent of time observed at Ethiopian primary schools. Especially within low- and middle-income countries, teacher-centred pedagogies are not only widespread but also often emphasise surface (rote) learning (Nag, 2014). Within classrooms, this often manifests itself in the form of students copying large amounts of notes from the blackboard and then practising what they copied through a series of individual seatwork exercises (e.g., Honeyman, 2015b). In Ethiopia, for example, Frost and Little (2014) found copying notes and individual seatwork to jointly account for around 28% of classroom time whereas classroom discussions and group work only accounted for around 4%. Both note copying and individual seatwork can be seen as indicators for teacher-centred classroom practices in their most "dominant" (Ngware et al., 2012) form. Especially prolonged periods of copying notes and individual seatwork exercises, indiscriminately administered to all students in class, may be perceived as boring and potentially meaningless by students ("autonomy-suppressing"; Assor et al., 2002). Insomuch as perceived autonomy can be associated with feelings of competence and favourable cost evaluations, suppressed autonomy through imposed prolonged activities such as note copying and individual seatwork may lead to the feeling of lack of competence and unfavourable cost evaluations ("a-motivation"; Assor et al., 2005). Both teacher-dominant classroom activities may also imply stifling students' active participation in classrooms. To the extent that active participation is indeed a way to promote task value amongst students (Urdan & Turner, 2005), copying notes and individual seatwork should thus also entail negative effects on perceived daily utility of lesson content.

Teacher-centred approaches towards learning may not only be associated with negative motivational effects. Especially within the direct-instruction paradigm (P. A. Kirschner et al., 2006), the importance of teacher explanations has been emphasised. Whereas classroom discussions and group work are bi- or multidirectional interactions, teacher explanations are "one-way teacher-to-learner transmission" (Schweisfurth, 2013, p. 147). Here, teachers use clear guided explanations that carefully walk students through new content and help them to internalise the key learning points (Rupley et al., 2009; Winograd & Chou, 1988). In this respect, teacher explanations provide students with a structure that is "competence supportive" (Vansteenkiste et al., 2012, p. 432). Ideally, teacher explanations provide students with guidance and directions that are aligned with their needs and problems during the learning process. This in turn may lead to more favourable cost and potentially daily-utility perceptions of learning content. Hence, teacher-centred pedagogies in the form of copying notes and individual seatwork may have negative motivational effects whereas teacher-centred activities in the form of teacher explanations may have positive effects on students' competence, utility, and cost perceptions.

A final comment is warranted. Since classroom activities constitute the concrete learning opportunities that students are expected to engage with in a specific moment, they are best measured repeatedly during a lesson. Based on these situation-specific ratings of classroom activities, however, a measure of the classroom climate of the corresponding lesson can be obtained. Through aggregating the individual ratings, the allocation of lesson time across student- and teacher-centred activities can be determined. These aggregations could then be considered as measures of the lesson-specific classroom climate (Marsh et al., 2012; Morin et al., 2014).

2.2.5. Covariates of achievement motivation

In addition to the different criterion and predictor variables, a set of covariates were included in this present DPhil project, which were assumed to be associated with different outcome variables within the nomological network around achievement motivation in Rwanda. Covariates controlled for included age, gender, and school type. First, age was controlled for during the analysis. Previous studies have shown that students' self-concepts (Gaspard et al., 2020; Marsh, 1989; Orth et al., 2020) as well as value perceptions (Gaspard et al., 2020) decrease with age (for a theoretical discussion explaining these results, see, for example, Harter, 2006). Möller et al. (2020) found age to moderate the relations between achievement and subsequent self-concept within I/E models. Second, gender was controlled for. After all, Eccles' et al. (1983) first iteration of the modern expectancy-value theory was developed to explain gender differences within career choices. There is also increasing evidence that despite general gender similarities in academic performance (Halpern et al., 2007) boys and girls differ in terms of their self-concept (Marsh, 1989; Niepel et al., 2019) and value perceptions (e.g., Eccles et al., 1998; Eccles, Wigfield, et al., 1993). For example, as part of their meta-analysis of gender differences amongst a set of different mathematics-related outcome variables, Else-Quest and colleagues (2010b) estimated gender differences in mathematics performance as well as mathematics-related anxiety (i.e., a cost-related measure) using the 2003 wave of the Programme for International Student Assessment (PISA) survey. The sets provided data on mathematics-related performance, beliefs (i.e., mathematics self-concepts and self-efficacy) and anxiety ("I get very nervous doing mathematics problems") amongst children aged 15. Using Cohen's *d*, they found that across all 35 countries (none of which were in Africa) for whom data was available, the weighted effect size of the gender gap in mathematics anxiety was significant and in favour of boys ($d = -0.28, p < .01$) whereas gender differences in mathematics achievement, which again favoured boys, were significant but negligible. Third, school type was controlled for. In many parts of the world, ability grouping of students is common. Students can be grouped at different levels within a country's school system (e.g., class-level or school-level). Probably the most common form of ability grouping is placing students into school types based on their performance during (national) examinations (i.e., explicit school-level tracking; Trautwein, Lüdtke, Marsh, et al., 2006), for example at the end of primary school. The underlying rationale for systems of "selective secondary education" (Ireson & Hallam, 2001, p. 2) has been the implicit assumption that teaching can be better tailored to the needs of students if they are grouped based on their ability levels. Critics, however, argued that being sorted into a low-achievement track may have negative consequences for students' motivational levels whereas being placed into high-achievement track may promote one's achievement motivation (e.g., Oakes, 1985). Even though a number of students documented that motivational effects associated within ability tracks appear not to be strong (e.g., Marsh et al., 2000; Preckel & Brüll, 2010; Trautwein, Lüdtke, Marsh, et al., 2006), some studies did indeed find systematic differences between students within low- and high-ability tracks in terms of their motivational levels (Wolff et al., 2021), at least initially (Liu et al., 2004). Hence, controlling for school type as a proxy for low- and high-ability tracks appears to be a warranted especially within novel research contexts.

2.3. Achievement motivation as a culturally dependent construct

Within the context of this DPhil, the case is made for motivational research in diverse cultural settings. It reflects the postulate that achievement motivation is a cultural concept. Already in 1983, Eccles and colleagues identified the "cultural milieu" (p. 80) as a determinant of students' disposition to interpret past events, and to set personal goals, which in turn are assumed to shape their expectancy and value beliefs. This implies the possibility that differences within cultural milieus can express themselves through differences within the way motivational dimensions manifest themselves. In fact, Eccles et al. (2004, p. 169) found that "cultural differences in a wide array of activity and behavioural choices, particularly in the achievement domain, reflect cultural differences in success expectations and subjective task value-related beliefs, which, in turn, likely result from cultural differences in the wide range of social experiences that shape human development" (Eccles et al., 2004, p. 169). This implies the conviction that human behaviours vary across cultures and these variations within human behaviours are related to the cultural environment. As a result, Eccles et al. (2020) have increasingly emphasised that modern expectancy value theory is not only "situationally specific”, which refers to a conceptualisation of expectancy and value beliefs that includes the momentary level, too, but also “culturally bound” (p. 2)

Why would culture shape achievement motivation? Culture can be defined as “any kind of information that is acquired from other members of one’s species through social learning that is capable of affecting an individual’s behaviours” (Heine, 2016, p. 5; based on Richerson & Boyd, 2005). In other words, culture comprises of those intersubjective ideas, beliefs, habits, and values that individuals acquire from (i.e., socialisation) and eventually share with their social environment (e.g., institutions, norms, practices, and artifacts; Markus & Kitayama, 2010). It is through this shared knowledge that culture helps to address the “information gap” (Geertz, 1973, p. 50) that exists between what is hardwired into the human brain and what is needed to survive. The information culture provides can largely be divided into two categories. First, it provides members with some guidance about how to obtain valued results in an otherwise uncertain social environment (i.e., “procedural knowledge”). Second, it provides ideas on what should be considered true and false (i.e., “declarative knowledge”) in a world that is otherwise void of absolute truths. The latter shapes the way individuals describe and evaluate social objects such as contexts (“event representation”), people within contexts (i.e., “person representation”), and behaviours of people within contexts (“norm representation”). Person representations concern evaluations not only of social groups and other people but also of one’s own self (C. Chiu & Hong, 2007, pp. 789–794). Self-representations are defined as “cognitive generalizations about the self, derived from past experience, that organize and guide the processing of self-related information contained in the individual's social experiences” (Markus, 1977, p. 64). Culture is therefore not just around the self. It is “intrinsic to the self” (Heine, 2001, p. 884). Motivational (self-) evaluations represent generalisations about the self within typical achievement situations. Hence, culture is intrinsic not only to the self as a whole but also to the self within typical achievement situations, which in turn implies that culture is also intrinsic to achievement motivation.

A number of examples highlight that culture may indeed shape achievement motivation. First, culture may moderate relations between competence beliefs and academic performance. Cross-cultural research found that compared to Western students, Asian students exhibit higher achievement especially in subjects such as mathematics or science (J. Lee, 2009; Wilkins, 2004) while exhibiting lower levels of self-concept (J. Lee, 2009) and self-efficacy (Klassen, 2004; Scholz et al., 2002), and higher levels of anxiety (J. Lee, 2009). One explanation for this result highlights that Western countries are mainly individualistic whereas East-Asian societies are predominantly collectivist in their cultural orientation (M.-T. Wang et al., 2019). Within collectivistic cultures, individuals are tightly assimilated into their in-groups such as family or community (Hofstede et al., 2010). A positive self-view is therefore determined especially through the social standing that is ascribed to oneself by others. By contrast, in individualistic societies, people feel loosely connected to their in-groups (Hofstede et al., 2010). A positive self-view is therefore less dependent on social obligations and more on personal achievement experiences. In this respect, individuals tend to self-enhance, which describes the tendency to exaggerate one's achievements while downplaying one's failures (e.g., Heine, 2003, 2005). This would be in line with a finding reported by Klassen (2004). He reviewed the literature on the culture/self-efficacy nexus and found that respective competence measures and achievement were related more strongly within collectivist contexts. He therefore concludes that "participants from the collectivist groups are inclined towards greater accuracy or 'realism' in predicting subsequent performance" (p. 225). If effects on achievement are mediated by engagement, then competence beliefs should also be stronger predictors of engagement especially in collectivist countries, something that was partly supported by Lei's et al. (2018) meta-analysis.

Second, culture may also shape value perceptions. Most cultural work focusses on values because “values, more than practices, are the stable element in culture” (Hofstede et al., 2010, p. 28). According to Eccles and colleagues (1983), utility value in particular reflects one’s personal goals one desires to attain. Initially, goal research focussed chiefly on the set of competence goals (e.g., Elliot & Murayama, 2008). This focus was then increasingly extended to include social goals(e.g., Dowson & McInerney, 2003; Urdan & Maehr, 1995; Wentzel, 1998). For example, based on their exploratory work in the Philippines, a collectivist country, King and McInerney (2019) found that student engagement is also driven by family support goals, which are about “supporting/assisting one’s family” (p. 342). Hence, to the extent that utility value is derived from one’s desired end states, the adoption of social goals should then also lead to the adoption of utility-value beliefs that reflect not only personal but also group needs (Eccles et al., 2004; Wigfield et al., 2016). However, motivation may not only vary in terms of its dimensional structure. Cross-country examinations of the association between utility value and achievement within mathematics revealed that correlations were more substantial in Western than in East Asian countries. It was partly negative in China and Taiwan, which suggests it was more important for low-performing Chinese and Taiwanese students (OECD, 2013b). Hence, culture may not only shape the manifestation of value beliefs but also their linkages with performance and thus engagement (M.-T. Wang et al., 2019).

Third, culture may also influence the level of verbal engagement exhibited in class. Verbal engagement of students is a cornerstone of student-centred pedagogies (e.g., Bremner, 2021). They facilitate dialogue especially through assigning more control over learning processes to students (Schweisfurth, 2013). At the classroom-level, this implies power structures that are more equally distributed between students and teachers (Hofstede et al., 2010). Even though this pedagogical style is generally associated with many learning benefits, its introduction within the Sub-Saharan region was met with mixed results (Schweisfurth, 2011). Failure in this respect was attributed, amongst others, to the hierarchical structure of African societies with its focus on dominance and subordination (Tabulawa, 2013), something that Hofstede et al. (2010) labelled “high power distance” (p. 60). Power distance at the classroom level often comes along with expectations on the side of teachers as well as students that teachers are predominantly the ones who control content and the process of learning (Hofstede et al., 2010). In Africa, this often manifests itself in the form of teachers teach and students listen (e.g., B. Fuller & Snyder, 1991; Stambach, 1994), which obviously contradicts many principles of student-centred pedagogies. To the extent that one does not allow for the time required for current classroom cultures to shift, one would expect students in Sub-Saharan Africa to be less verbally engaged than, for example, Western contexts that have generally been associated with student-centred pedagogies (Hofstede et al., 2010; Tabulawa, 2013).

Fourth, culture may influence students’ formation of motivational beliefs. Culture may shape the organisation of a school system (M.-T. Wang et al., 2019). Important school-related features are teaching formats and teacher evaluation practices. Within a unidimensional teaching format all students in class are taught in a standardised way whereas a multidimensional format implies learning content being adjusted to students based on their needs. Performance evaluative formats may vary in terms of their level of openness. Students may receive feedback on their performance in a private or public manner whereby the latter implies that the teacher’s assessment is shared with a student in front of the class (Karasawa et al., 1997; Rosenholtz & Rosenholtz, 1981). A unidimensional teaching structure coupled with public feedback should facilitate social comparisons whereas a multidimensional format with private feedback practices should hinder social comparisons (Rosenholtz & Rosenholtz, 1981). The effects of the school system on students’ motivational belief-formation have also been observed within research on the I/E model. Marsh, Abduljabbar, et al. (2015) performed a cross-country test of the I/E model involving Western, Asian, and Middle Eastern, Islamic cultures. Amongst Middle Eastern students, cross-domain effects of achievement on competence were close to zero while within-domain effects were significantly smaller than for Western and Asian students. Marsh and colleagues explain this finding also through the local cultural context: “Middle Eastern Islamic students do not receive as much evaluative feedback about their achievement as do Western and Asian students and are not socialized in such a way as to critically evaluate their academic skills in relation to classmates” (p. 192).

Fifth, the perceived similarity of school domains may be influenced by culture as well. This may be true especially for those domains that researchers consider to be in close proximity otherwise. Marsh and Hau (2004) tested the generalisability of the I/E model between mathematics and the verbal domain across 26 countries. For each of the 26 countries they also estimated cross-domain path coefficients for mathematics achievement to verbal self-concept and verbal achievement to mathematics self-concept. Even though not all were statistically significant, the 52 coefficients were generally negative. These contrast effects suggests that irrespective of the county context, mathematics and verbal domains are generally perceived to be different. Chiu (2008) tested the I/E model across 28 countries using mathematics and science, two domains that are considered to be similar (e.g., Marsh, Lüdtke, et al., 2015). According to dimensional theory, similarity between domains should lead to assimilation effects between both subject areas (Möller & Marsh, 2013). Chiu, however, found only 24 out of 56 coefficients to be positive. Based on these results, Chiu also grouped countries. Significantly more countries (10 out of 28) exhibited contrast effects across both domains than showing assimilation effects (4 out of 28). In the remaining 14 cases, countries showed both assimilation and contrast effects even though not all of them were very significantly different from zero. Even though Chiu’s data set did not allow to isolate the actual drivers behind those country differences, she found “a trend in that most integrating countries are located in the Middle East and that most contrasting countries use English as part of their official language(s)” (p. 250). Also, based on a cross-country analysis, Else-Quest et al. (2010a) found that equal access to education between boys and girls predicted gender parity within mathematics-related competence beliefs, intrinsic and utility value, and achievement. These results suggest that cultural norms around the role of males and females also express themselves through motivational processes (Eccles et al., 1983).

Sixth, even though there has not been much cross-cultural research on the intrapersonal variability within achievement motivation, the scarce evidence available suggests that the size of intracluster correlations (ICCs) may vary across country contexts. Examining ICCs for expectancy and utility value within the science domain, Salmela-Aro et al. (2021) found them to be lower in the Finnish (.19/.27) than in the Chilean (.43/.52) context. Hence, in the case of Finish students, around 73 to 81 per cent of variability within motivational states is due to differences between situational factors rather than between person differences whereas it is about 48 to 57 per cent in the case of Chilean students. This study alone may not suffice to conclude that ICCs are moderated by culture. However, the study warrants further investigations into whether or not ICCs are higher in non-Western contexts and why. Salmela-Aro et al. (2021) also found that expectancy and value states predict engagement. However, perceived challenge, a cost measure, predicted disengagement in Finland but not in Chile. The authors explain the results through “the characteristics of the Chilean educational context where, in general, science classes are perceived as boring and not providing enough challenge” (p. 422). These examples highlight that culture shapes achievement motivation at least indirectly. This in turn bears the question of how best to control for culture within the study of achievement motivation. The next section outlines the two main approaches in this respect.

2.4. The universalist approach to studying motivation within cultures

One distinguishes between the emic and the etic approach towards the study of human psychology within a cultural context (e.g., Berry et al., 2002). Within research on cultural influence, both concepts describe different ways to determine meaning and operationalisation of psychological variables of interest for a given cultural context.[8] Etic research, which has been the dominant stance for some time especially in cross-cultural psychology (R. B. King et al., 2018; Zusho & Clayton, 2011), implies it is the scientist who defines and operationalises psychological constructs of interest. In this respect, their meaning and operationalisation is generally derived from a priori psychological theories that are assumed to be independent of any cultural framework (Berry et al., 2002; Chao & Lambert, 2013; F. M. Cheung et al., 2011). Etic research is therefore not concerned with the exploration of psychological phenomena locally but focusses instead on the effects of culture on psychological variables. Accordingly, etic researchers generally rely on cross-context studies to explore the relevance of culture for psychological processes (e.g., R. B. King et al., 2018). A main point of critique of the etic perspective has been that respective theories chosen to guide study design and operationalisation of constructs are often not free of cultural assumptions. It is especially current theories of mainstream psychology that frequently appear to be influenced by a Western cultural background whose implications are then tacitly generalised across all other cultures ("pseudoetics"; Triandis et al., 1971, p. 6). As the example of Marbell and Grolnick (2013) highlights, these tacit assumptions may not always

[8] One needs to distinguish emic/etic research approaches from emics/etics. Whereas emic/etic approaches refer to how the meaning and operationalisation of culturally sensitive concepts are determined (Berry et al., 2002; Chao & Lambert, 2013), emics/etics refer to the level of universalism of psychological phenomena. Emics refers to 'concepts that are unique to one culture' while etics 'are concepts that can be found in most, if not all, cultures' (Kitayama & Cohen, 2007).

hold. Marbell and Grolnick were interested in testing to what extent different parenting styles (i.e., “parental control” versus “autonomy support”) is related to children’s intrinsic and extrinsic motivation within the collective society of Ghana. In general, their research design is best described as etic since it is predominantly built upon scales that have been mainly validated in non-African contexts (e.g., USA). Within a first study using a sample of 95 six-graders (aged 12 on average), the authors wanted to examine some of the psychometric properties of these scales within the Ghanaian context. Not surprisingly, they found especially the scale on parenting not to be reliable. Additional qualitative explorations revealed that “some items that tapped into children’s experience of autonomy support in the US were interpreted as independence and perceived negatively by some Ghanaian adolescents” (p. 88).

Within the realm of emic research, one attempts to reduce cultural bias and to discover the actual meaning and relevance of psychological phenomena within the culture it appears by adopting a local perspective (Berry et al., 2002; Chao & Lambert, 2013; F. M. Cheung et al., 2011). Accordingly, psychological constructs are not derived from any “outside” theory whose validity may never have been established for a given cultural context. Instead, one aims at unpacking meaning of cultural concepts using local knowledge uncovered especially through qualitative methods involving those that live within the (novel) culture studied (Berry et al., 2002; Chao & Lambert, 2013; F. M. Cheung et al., 2011). Within subsequent studies, the internal and external nomological network around the locally grounded concepts can then be studied. Both the qualitative and quantitative results can then be compared with the mainstream (i.e., Western-borne) literature, which should thus provide an indication of cultural differences of the novel context being studied. As already mentioned, applying this logic to the case of achievement motivation amongst Rwandan students is a key aspect of the theoretical rationale of this DPhil project. In this respect, it may contribute to the establishment of universal theories on achievement motivation. Especially systematic reviews of results from single-culture studies, like the ones reported in this document, may allow to derive hypotheses and eventually theories that hold not only within a single culture but potentially universally across many different cultures (e.g., Berry, 1969). This latter step implies a universalistic stance towards the study of motivation within cultural contexts. Such a stance postulates that motivation is a universal process that is inherent to all humans but that the way motivation manifests itself may vary across cultures (e.g., Zusho & Clayton, 2011). Universalistic studies to qualitatively explore locally grounded phenomena often rely on concepts derived from the Western mainstream literature ("imposed etics"; Berry, 1969, 1990). For example, Clayton and Zusho (2016) mapped out those cognitive affective factors that motivate Jamaican students for their academic work at university. Within a qualitative study using open-ended questionnaires, they asked undergraduates "What motivates you (and others) to persist in school?" or "Why are some students unmotivated in school?" (p. 12). In this respect, the concept of motivation represents an entry point that allows the authors to embark on their qualitative exploration of the factors that drive engagement. Clayton and Zusho did not explore whether or not students exhibit any motivation but focussed only on the way it materialises itself in the Jamaican context.

Qualitative studies within universalistic research may not only help to detect novel aspects within the local manifestation of achievement motivation. They may also provide important raw data that may be of use to derive those self-report measures needed to eventually study the nomological network making up achievement motivation. Within universalistic research, the use of qualitative methods to ensure the validity of quantitative measurement instruments is highly encouraged because "there is a danger in assuming that an instrument developed in one culture will transfer to another culture" (Zusho & Clayton, 2011, p. 245). As the case of Marbell and Grolnick (2013) shows, this concern is not unwarranted.

Employing qualitative methods to ensure the validity of measurement instruments is in line with the bottom-up approach towards scale development. Scale development describes the processes researchers engage with to obtain those (self-report) items they need to operationalise a (latent) construct (e.g., DeVellis, 2017; McCoach et al., 2014). A top-down ("deductive") scale-development approach implies researchers either deriving scales from the literature or from their own understanding of the theoretical concept being studied. As the example of Marbell and Grolnick (2013) shows, this may fail especially in "novel" research contexts. In the case of bottom-up ("inductive") approaches, scale developers consult members of the population of interest about their association with the concept under consideration. These answers will then be grouped into different themes. Each theme may then reflect different aspects of the latent construct, thus items within each thematic group may help to operationalise respective aspects (Hinkin, 1995, 1998). The inductive item-generation is predominantly helpful in cases where there is not much theory to guide the scale development process around a construct of interest (Hinkin, 1998). For example, King and McInerney (2019) explored different goals that Filipino secondary students pursued through their studies by administering the open-ended question "Each student has a different reason for studying. What are your reasons for studying?" (p. 341/ 342). The qualitative analysis of the answers provided suggest that the majority of goals specified by students could be classified with a conventional (etic) achievement-goal framework (Elliot & Murayama, 2008). In addition, around 15 per cent of the answers provided indicated goals related to family support (e.g. "I want to help my family financially in the future.", R. B. King & McInerney, 2019,; see supplementary material). Those statements were then used to generate items for a scale to be used to measure family-support goals quantitatively ("I want to do well in school so that I can help my family financially in the future", p. 343). In follow-up studies, this scale was then shown to be predictive of students' self-reported learning engagement.

To sum up, emic research processes are important complements to etic research undertakings (R. B. King & McInerney, 2014, 2016). They help to explore local manifestations of psychological processes such as motivation in cultural areas that have not been previously subject to empirical research. In this respect, they may not necessarily contradict etic premises. However, they may complement etic theories around psychological processes by highlighting important latent aspects that have been previously ignored. This in turn may also lead to endeavours to test new hypotheses involving these novel latent constructs quantitatively. Here, emic research, as part of an inductive item-generation process, helps to operationalise these constructs through scales that are indeed locally grounded. It is for these reasons that cross-cultural researchers such as John W. Berry argue that sound and valid etic knowledge around human psychology can only be attained on the back of well-designed emic studies into the psychological processes within different cultural contexts ("derived etics"; Berry, 1969, 1990).

2.5. Achievement motivation in African contexts

After having described how the internal and external nomological network of achievement motivation may be shaped by culture (section 2.3) and how to study achievement motivation within a cultural setting (section 2.4), this section briefly outlines the research gaps in the internal and external nomological network of achievement motivation within the African context. It provides additional justification of the study of achievement motivation within Sub-Saharan Africa, since achievement motivation as a concept has not yet been systematically studied there, at least from an educational psychology perspective. Different theoretical lenses have been applied to the study of education within Africa. Some of them touch upon motivational relevant aspects as well. There may be too many to provide a systematic review here. The focus is therefore put especially on those aspects of achievement motivation relevant to the present DPhil project. It is also important to point out that it would not be scientifically sound to speak of Africa as one context. Africa is a continent with a multitude of different peoples, religions, languages, historical parts, potential futures, and cultures. It is therefore better to speak of African contexts of which Rwanda is one. This overview here is structured along the elements of the internal and external nomological network outlined in sections 2.1 and 2.2. However, it starts with a brief overview of the underrepresentation of African contexts within “mainstream” motivational research.

The underrepresentation of Africa within mainstream motivational research is best illustrated through the examination of geographical distributions of studies that emerged within systematic reviews of motivational constructs of interest. Some examples are provided below. A systematic review aims to extract, synthesise, and evaluate results on a particular subject matter from relevant studies (Higgins, 2019). Within systematic reviews, it is common to highlight the geographical distribution of these published studies, which in turn can be taken as a yardstick by which to judge the geographical spread of research within a given domain across the different regions of the world. A meta-analysis of the development of domain-specific self-concept found “no African” (Orth et al., 2020, p. 151) samples amongst the 143 published longitudinal studies reviewed. Within a meta-analysis (Kriegbaum et al., 2018) on the importance of motivation as a predictor of school achievement relative to intelligence, 77 samples were reviewed. Only one was from Africa (i.e., Nigeria). Lei et al.’s (2018) meta-analysis on the relationships between engagement and academic achievement included 69 studies from geographical areas coded as Eastern (e.g., China or Korea) or Western (e.g., Europe). In addition, they also included a miscellaneous category that included a small number of studies from other geographical areas such as Africa. Roorda et al.’s

(2011) meta-analysis on the effects of student-teacher relationships on engagement found 99 independent studies (i.e., South Africa; Mboya, 1995). Cornelius-White's (2007) meta-analysis on the effects of student-centred teacher-student relationships reviewed 119 studies. None of them were from Africa. A welcome exception is the systematic review on the outcomes of learner-centred pedagogy within low- and middle-income countries reported by Bremner et al. (2022). Amongst the 62 journal articles reviewed, 29 were from African contexts. By contrast, Möller et al.'s (2009) meta-analysis on the classic I/E model across mathematics and the verbal domain included 69 samples from 66 different countries. None of them were from Africa. Möller et al. (2020) expanded their analysis by including other school domains as well. Amongst the 121 investigations they reviewed, only the 28-country study reported Chiu's (2008) included samples from Africa (i.e., Botswana, Ghana, and South Africa). As these examples highlight, research on the internal and external elements of the nomological network of achievement motivation carried out in Africa, probably with the sole exception classroom activities, is sparse at most.

But what does the available evidence suggest? Intrapersonal investigations of achievement motivation amongst African students that employ an intensive-longitudinal design seem to be largely missing. Studies on the person-specific structure of achievement motivation within African contexts exist. According to Sini et al. (2018), constructs similar to value and expectancy beliefs emerged in an exploratory factor analysis of data also collected in Kenya. Moreover, value beliefs may not necessarily be a unidimensional construct. McCoy et al. (2014) found that approximately half of students in their sample from Ghana were more intrinsically motivated. The other half were predominantly extrinsically motivated (e.g., "I try to do well in school because I might get a reward if I do well"), an aspect hardly discussed within expectancy-value theories (Gaspard et al., 2015). However, a major limitation of these studies was the reliance on closed-ended, preconceived measurement instruments derived from the Western literature. This may have prevented the incorporation of genuinely African perspectives (R. B. King et al., 2018). Whereas the number of studies that explore the dimensional structure of achievement motivation within Western classrooms slightly increase (e.g., Dobie, 2019; Erickson, 2019), respective investigations within African contexts continue to be lacking.

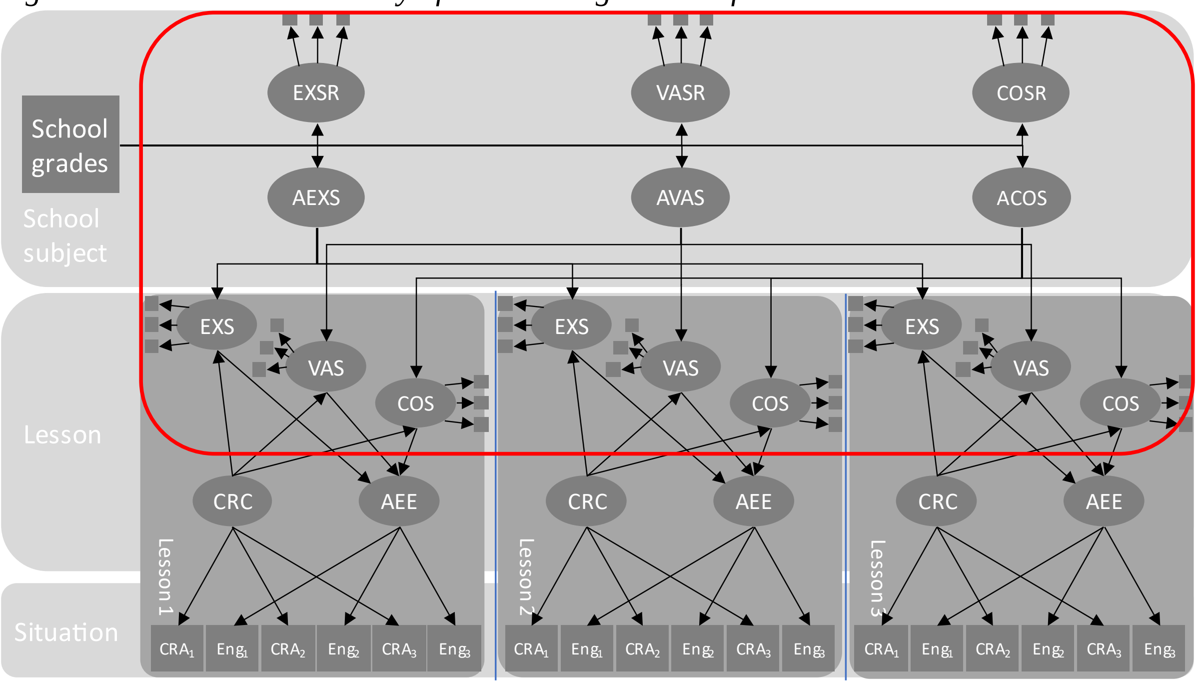
Qualitative studies within African contexts seem to be about general (self-) perceptions students attach to schooling as a whole. A common interest observed in this respect is about "why individual children […] do go to school" (Marshall, 2016, p. 423; Posti-Ahokas & Palojoki, 2014). This focus is rather different to what is usually discussed within expectancy-value theory, which is generally about explaining students' achievement behaviours and choices within schools (e.g., Eccles et al., 1983). Two concepts explored within the qualitative literature on motivation in Africa are aspiration and agency (e.g., DeJaeghere, 2018; Marshall, 2016; Posti-Ahokas & Palojoki, 2014; Tafere, 2014). The latter implies taking those necessary steps to attain a valued outcome (DeJaeghere, 2018) and has been found to be associated with school attendance (DeJaeghere, 2018). As a construct, it is related with expectancy beliefs in that self-beliefs are the key to developing a sense of agency in the first place (Bandura, 1986). According to Appadurai (2004), aspiration is related to "wants, preferences, choices, and calculations" (p.67). From an educational-psychological viewpoint, the concept can be interpreted as representing achievement goals, which constitutes those "desired end states" (Hulleman, Schrager, et al., 2010, p. 423) individuals adopt in education (DeJaeghere, 2018; Marshall, 2016; Posti-Ahokas & Palojoki, 2014; Warrington & Kiragu, 2012) and other life domains. In this respect, social-status goals seem to be common reasons for going to school. DeJaeghere (2018, pp. 245–248), for example, reports of a Tanzanian girl who entered a vocational school because she aspired to "an average life as other people are living […] unlike the life of beggars". At school, she then assessed if school subjects "helped" her to attain that goal.

In contrast to studies on the internal structure of achievement motivation, investigations of external nomological net appear to be more common. Studies that use achievement motivation as the criterion variable are reported. In Tanzania, Sakata (2022) carried out classroom observations during 17 lessons and collected data on student attitudes towards learning including confidence and motivation, the latter being a proxy for value perceptions. She did not find classroom activities associated with student-centred pedagogies to predict any of those (self-) evaluations. By contrast, Mboya (1995) looked at teacher characteristics such as the extent to which they provide structure or are emotionally involved with their students and found these characteristics to predict academic self-concept. As part of her 28-country investigation of GI/E model relations, the above-mentioned study by Chiu (2008) included Botswana, Ghana, as well as South Africa. Using TIMSS data, she examined relations between achievement and expectancies (academic self-concept) across mathematics and science. Across the three countries, prior achievement positively predicted expectancy beliefs within domains. As for effects between domains, there was some variation. In Ghana, Chiu found assimilation effects whereas in Botswana and South Africa she detected contrast (i.e., negative) effects, which suggests country context to be a moderating variable. As for the outward-facing part of achievement motivation's nomological network, studies in Ghana found self-concept beliefs but also intrinsic value to predict educational outcomes such as aspiration and engagement (Ansong et al., 2019; Marbell & Grolnick, 2013). By contrast, studies on the nexus between classroom activities and engagement does not seem to have been studied within African contexts (compare, for example, Bremner et al., 2022). To sum up, there is evidence on the internal structure of and external relations around achievement motivation within African contexts. However, the evidence appears to be unsystematic in the sense that it does not allow to map out both the internal and external nomological network of achievement motivation. This is particularly troublesome because Africa is best understood as a geographical cluster of many different cultural, national, and social contexts. This research gap in turn renders any evidence base upon which to develop truly universal theories of achievement motivation incomplete. This DPhil project therefore provides a first attempt to map out achievement motivation's nomological network within one of Africa's many contexts: Rwanda.

3. Summary and research questions

Within the present DPhil project, the nomological network of achievement motivation within the Rwandan school context will be investigated. Figure 4 on the next page depicts the aspects of the nomological network of the concept of achievement motivation tested in Rwanda. The aspects were derived from the review of the motivational literature outlined in the previous chapter. The schematised nomological network is depicted hierarchically consisting of the three situation-specific levels, which is nested within the lesson-specific level, which in turn is nested within the school subject-specific level. The case depicted in figure 4 consists of three situations within each of the three lessons, which all belong to the same subject. The complex concept of motivation is demarcated from other elements of the network through the red lines. As the figure highlights, motivation is hypothesised to be a multilevel and multidimensional concept (red rectangle). It is multidimensional in that it consists of the following motivational factors: expectancy (EXS/AEXS/EXSR), value (VAS/AVAS/VASR), and cost (COS/ACOS/COSR) (self-)evaluations. It is multilevel in that it spans the lesson-specific as well as the school subject-specific level. Motivation at the school subject-specific level is conceptualised as both school subject-specific motivational self-reports (EXSR/VASR/COSR) as well as average motivational states (AEXS/AVAS/ACOS). The former is gauged directly through self-report items where the latter is inferred from lesson-specific motivational states, which in turn are measured through self-reports (EXS/VAS/ COS). The red rectangle graphically defines the internal aspects of the nomological net studied within the Rwandan context. Elements outside the rectangle constitute the external elements of the nomological net of achievement motivation. (Prior) school grades are depicted as a school subject-specific predictor variable that is assumed to determine both motivational self-reports as well as school subject-specific average motivational states. At the lesson-specific level, classroom climate (CRC) is assumed to predict lesson-specific expectancy (EXS), value (VAS), and cost states (COS). The classroom climate is the average of all situation-specific ratings of classroom activities observed (CRA_1 to CRA_3) taken within each of the three lessons. Within the model tested, the only predicted external variable is engagement. As with classroom activities, it is a situation-specific construct nested within lessons. The lesson-specific average engagement exhibited (AEE) is the mean of all situation-specific engagement observations taken within a lesson (Eng_1 to Eng_3).

Figure 4: Schematised summary of the nomological net of motivation tested in Rwanda



Note. EXS, VAS, and COS stand for lesson-specific expectancy, value, and cost states. AEXS, AVAS, and ACOS stand for school subject-specific average expectancy, value, and cost states. EXSR, VASR, and COSR stand for school subject-specific expectancy, value, and cost self-reports. CRA_i stands for classroom activities observed at situation I, i = 1…3. Eng_i stands for engagement observed at situation I, i = 1…3. Circles represent latent constructs whereas boxes constitute manifest indicators. Within levels, motivational factors may be correlated. To avoid cluttering, this is not depicted.

The nomological net within and around achievement motivation tested in Rwanda was derived from the (Western-borne) motivational literature. The way latent constructs within the nomological net (and engagement) were operationalised empirically was not derived from the academic literature. In line with the universalist approach to the study of psychological phenomena within cultural contexts, it is assumed that achievement motivation is a driver of behaviours of Rwandan students but that the way it manifests itself in Rwanda cannot be fully derived from the Western-borne motivational literature. The project therefore consisted of an exploratory-qualitative part to scope out the level of multidimensionality of the internal structure of achievement motivation as well as a quantitative-confirmatory part to test a set of hypotheses around achievement motivation that were outlined in the previous literature review. That way, this DPhil project will provide a first systematic exploration of both the internal as well as external aspects of the nomological net associated with achievement motivation in Rwanda. The exploratory and the confirmatory part of the project will be linked up through a bottom-up (i.e., emic) scale development process through which qualitative insights on the dimensional structure of achievement motivation in Rwanda will be translated into self-report scales to measure relations quantitatively involving the concept. Based on the literature review presented above, the DPhil project will be guided a set of five research questions, each with their respective sub-questions. These research questions are arranged into the following five thematic blocks:

1. **What is the dimensional structure of achievement motivation amongst Rwandan secondary students?**
   1.1. *Can the different dimensions of the motivational experiences gauged from Rwandan secondary students qualitatively be adequately described through the different cognitive-emotional facets proposed by the expectancy value theory?*
2. **To what extent does prior academic achievement in the form of school grades predict subsequent motivational beliefs amongst Rwandan secondary students?**

2.1. Are relations between prior achievement and motivational beliefs significant and positive within domains?

In line with the GI/E model (e.g., Möller et al., 2016), it was predicted that prior achievement would significantly and positively predict subsequent motivational beliefs. As further explained in manuscript two, it was assumed that respective effects are the strongest in the case of competence beliefs and the weakest in the case of utility beliefs.

2.2. Does academic achievement exert negative cross-subject effects on motivational beliefs between two subjects that are both considered to be different?

In line with the dimensional comparison theory (Möller & Marsh, 2013) associated with the GI/E model (e.g., Möller et al., 2016), it was predicted that prior achievement in one subject (e.g., Mathematics) would significantly and negatively predict subsequent motivational beliefs in another subject that is considered to be rather different (e.g., English). As further explained in manuscript two, it was assumed that respective effects are the strongest in the case of competence beliefs and the weakest in the case of utility beliefs.

2.3. Does academic achievement exert positive cross-subject effects on motivational beliefs between two subjects that are both considered to be part of the quantitative school domain?

In line with the dimensional comparison theory (Möller & Marsh, 2013) associated with the GI/E model (e.g., Möller et al., 2016), it was predicted that prior achievement in one subject (e.g., Mathematics) would positively but weakly predict subsequent motivational beliefs in another subject that is considered to be rather similar (e.g., Chemistry).

2.4. *Does academic achievement exert positive cross-subject effects on motivational beliefs between two subjects that are both considered to be part of the language domain?*

In line with the reasoning put forward by van der Westhuizen et al. (2022), it was assumed that comparing one's own abilities within languages that are all well integrated into one's lives will not lead to assimilation effects but instead to contrast effects. Thus, it was predicted that prior achievement in one language should weakly but positively predict motivational beliefs in another language in case of Rwandan day schools but negatively in the case of Rwandan boarding schools.

2.5. *To what degree are relations between academic achievement and motivational beliefs between and within school subjects moderated by school type?*

Due to the ability grouping within in the Rwandan education system, boarding schools are considered as 'schools of excellence' (Williams, 2019, p. 648) whereas basic-education schools are seen to be "for those who failed" (Williams, 2019, p. 646). It was expected that boarding-school students would exhibit higher prior school grades. Due to the reputational differences between basic-education and boarding schools, it was also expected that boarding-school students would hold more favourable levels of expectancy and cost beliefs across the four school subjects compared to basic-education students.

3. **To what extent does prior academic achievement in the form of school grades predict subsequent motivational states reported by Rwandan secondary students?**

3.1. *What is the level of school subject-specificity of average motivational states observed amongst Rwandan secondary students?*

It was assumed that average competence and cost states exhibit low levels of subject specificity. In line with Bong (1998), it was expected that both average competence and cost states exhibit substantial cross-subject correlations.

3.2. *What level of convergent validity is observed between motivational beliefs as well as average motivational states?*

Only moderate convergent validity between average competence and cost states on the one side and school subject-specific competence and cost self-reports on the other side was expected. It was expected respective correlations to be substantially below .75 (Gaspard et al., 2018).

3.3. *Does academic achievement exert positive effects on average motivational states within subjects?*

In line with the expected low subject specificity and convergent validity, it was predicted the GI/E model to only partly explain average competence and cost states. It was not expected that the internal frame-of-reference effects to apply to the average competence and cost states (Bong, 1998) manifested through regression coefficients that are not significantly different from zero.

3.4. *Does academic achievement exert negative cross-subject effects on average motivational states between two subjects that are considered to be different?*

Even though the GI/E model may not predict well average competence and cost states, average competence and cost states we expected to be predicted by the prior school grades within the corresponding subject.

3.5. *To what extent are relations between academic achievement and average motivational states moderated by school type, age, and gender?*

Different covariates were expected to predict average competence and cost states. It was expected that school type (i.e., basic education versus boarding schools), gender, and age would influence the levels of average competence and cost states.

4. **To what extent do classroom activities associated with student-centred and teacher-centred methods predict motivational states reported by Rwandan secondary students?**

4.1. To what extent are the student-centred and teacher-centred lesson activities used within Rwandan classrooms?

In line with the studies reported by Sakata (2022) and Frost and Little (2014), we expected teacher-centred activities to account for the biggest part of lesson time observed at Rwandan schools. Reflecting especially the anecdotal evidence provided by van de Kuilen et al. (2020), however, it was expected that the share of lesson time to fall on teacher-centred activities to be less than those 75 to 81% that Sakata (2022) and Frost and Little (2014) reported.

4.2. To what extent does the time allocated across student-centred and teacher-centred classroom activities predict Rwandan students' lesson-specific expectancy, utility value, and cost (self-) evaluations?

Within the present study, student-centred pedagogy was operationalized through the allocation of time to classroom discussions and group work. It was hypothesized that their use positively predicts students' lesson-specific expectancy, utility, and cost (self-) perceptions. By contrast, classroom activities in form of note taking and individual seatwork were taken as instances of a dominant teaching style (Ngware et al., 2012) and assumed to negatively predict expectancy, utility, and cost (self-) perceptions. Finally, even though teacher explanations are teacher-centred (but not teacher-dominant) practices, it was assumed that the time allocated to teacher instructions would exert positive effects on all three motivational states.

4.3. *To what extent are Rwandan students' experiences of expectancy, daily utility, and cost predicted by age and gender?*

It was expected that both age and gender partly influence expectancy, daily utility, and cost (self-) evaluations of students. To be more precise, it was expected older students to have lower competence beliefs and more negative cost beliefs and female students to have lower competence beliefs, higher utility perceptions, and more negative cost beliefs.

5. **To what extent do classroom activities associated with student-centred and teacher-centred methods as well as motivational states predict engagement levels that Rwandan secondary students exhibit during class?**

5.1. *To what extent is Rwandan students' behavioural engagement subject to intrapersonal variability both within and between lessons?*

In line with the literature, it was expected that behavioural engagement to exhibit substantial variability not only between students but also between lessons as well as within lessons. Moreover, it was expected that the biggest share of variability to originate at the within level.

5.2. *To what extent do classroom activities predict Rwandan students' behavioural engagement in class?*

Being a "situational" variable, it was expected that behavioural engagement exhibited during lessons to be influenced by contextual factors. Within lessons, it was considered that classroom activities to be major sources of contextual influences on engagement. In line with the literature, it was expected that those activities associated with student-centred practices to exert positive influences on engagement while those activities we linked with teacher-dominant practices were expected to exert negative effects. Teacher explanations were also assumed to be positively related to engagement levels.

5.3. *To what extent is Rwandan students' behavioural engagement in class related to their lesson-specific motivational self-reports?*

Being the "outward manifestation of motivation", it was assumed that engagement would be positively predicted by students' competence, utility, and cost self-reports. It was expected that competence but also cost (self-) evaluations would be stronger predictors of engagement than utility evaluations.

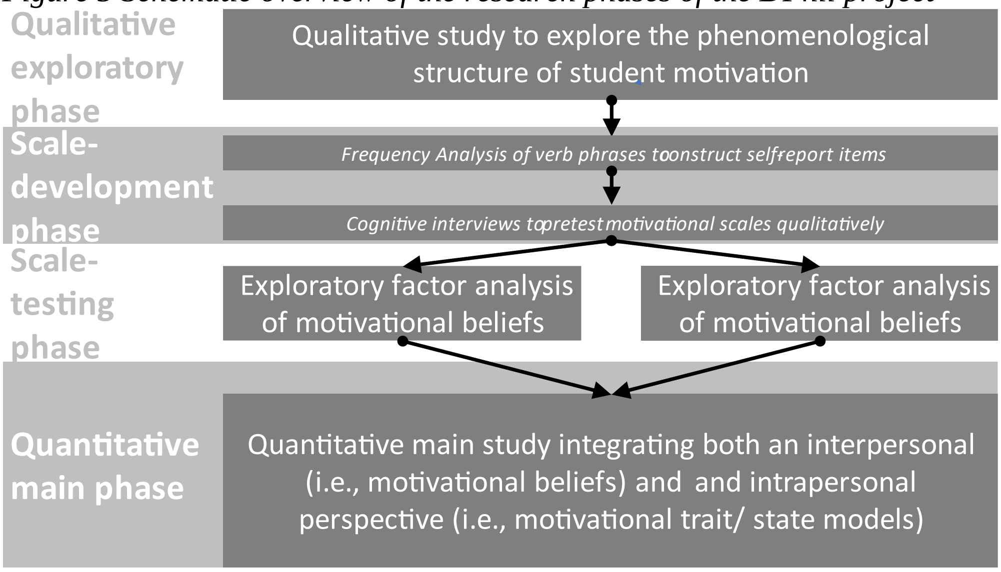
4. The methodology

Given the research implications stemming from the universalist approach chosen and given that motivational scales have never been developed for the Rwandan school context, the present DPhil project was implemented through three phases (i.e., the qualitative exploratory phase, the scale-development phase, and the main phase). In section 4.1 of this chapter, the overall methodology of the project as it cuts across the different research phases is presented. Additional details of the method chosen to implement the qualitative exploratory phase are spelt out in manuscript 1 in sub-section 5.1 of the results section. The outline of the scale-development efforts is normally part of the supplemental material of manuscript two. Within this DPhil document, it is included as sub-section 5.2 in the results section. The project's main phase is presented to the reader in the form of manuscript two to five. None of the manuscripts provide a full overview of the methods chosen to implement the main phase in its entirety. In fact, the studies presented by each of the manuscripts vary in terms of the school and student samples they use, the data-collection tools they employ, and the motivational constructs they target. To help the reader obtain a full and coherent overview of the main phase, section 4.2. spells out its key methods, overall sample composition, data-collection approaches, and key analyses. Section 4.3 outlines the general measures taken to ensure data quality throughout all project phases. Section 4.4 concludes the methodology chapter with a discussion of the ethical considerations that guided project design and implementation.

4.1. The overall project methodology: substantive, structural, and external

The overall research sequence of this DPhil project is summarised in figure five on the next age. The objective of the present DPhil project was to explore both internal and external aspects of the nomological network of achievement motivation amongst Rwandan students following a universalist approach to the study of psychological concepts within cultural contexts. In line with the universalist approach to the study of psychological phenomena within cultural contexts, the dimensional structure of achievement motivation was explored using qualitative methods during phase 1. Their flexibility makes qualitative methods particularly well-suited for exploring local perceptions around motivation within student descriptions (Bogdan & Biklen, 2007; Creswell, 2007; Merriam & Tisdell, 2015). In particular, open-ended questionnaires were employed, requiring students to describe their motivational associations. Emerging themes identified during the analysis of the questionnaire data collected during phase 1 were interpreted as dimensions that jointly make up the local manifestation of achievement motivation in Rwanda. Sub-section 5.1 of the results section provides a detailed outline of the methods employed during phase 1.

Figure 5 Schematic overview of the research phases of the DPhil project



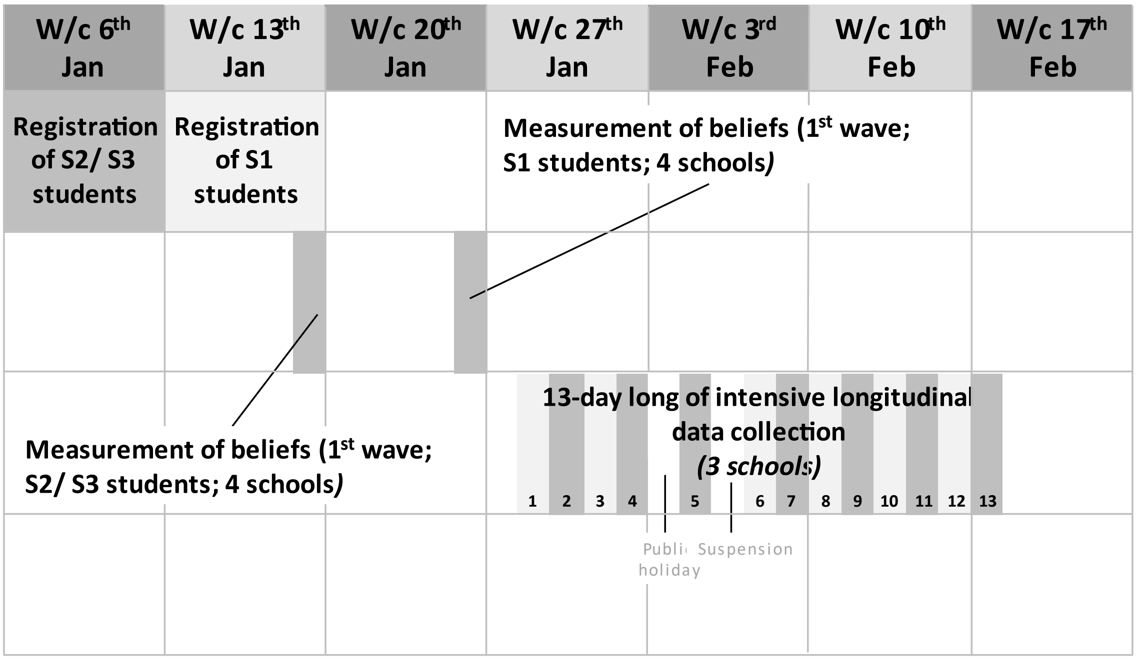
To avoid any of the pitfalls associated with a locally inappropriate conceptualisation and operationalisation of motivation (see above the discussion of Marbell & Grolnick, 2013), the measurement instruments required for the quantitative-confirmatory part of the project were developed next during phase 2. Phase 2 revolved around the development of both instruments to gauge motivation as well as engagement and classroom activities. Irrespective of the dimension or the level of specificity, achievement motivation constitutes person-internal processes and is thus commonly measured using self-reports (Appleton et al., 2006). To avoid common-method bias (Podsakoff et al., 2003), behavioural engagement was not measured through self-reports. Behavioural engagement refers to "the *observable* qualities of students' actions with academic tasks" (Skinner & Pitzer, 2012, p. 24; emphasis added). It was therefore decided to gauge behavioural engagement of students in class through observations by external raters (e.g., Ocumpaugh et al., 2015; Shapiro, 2011). The observation tool was then extended to include ratings of classroom activities (e.g., Stallings, 1977; World Bank, 2015). The observation tool was developed through a multi-staged approach consisting of a literature review to derive a provisional protocol, extensive field tests of the protocols, and incorporation of feedback by external raters. These stages are outlined in annex 5. The process implemented to obtain motivational self-report instruments involved substantial data collection and is thus outlined in section 5.2 of the results section. It consisted of a scale-development and a scale-testing phase. In line with universalist reasoning, the scale-development process was "inductive" (Hinkin, 1998). Accordingly, the themes that emerged during the analysis of phase 1 were used as a starting point to derive scales. Since each theme was assumed to represent a locally relevant dimension of motivation, each theme should have also ideally been represented by a separate scale. The raw data collected during phase one was of interest, too. The raw data consisted of descriptions of motivational experiences given by respondents in their first language Kinyarwanda. The data thus provided invaluable insights into how Rwandan students verbally express motivation (Fredricks et al., 2016; Gehlbach & Brinkworth, 2011). Verb phrases students used were therefore analysed and turned into self-report items during phase two. The scale-testing stage consisted of a series of sequentially related studies that helped to refine and condense the provisional self-report scales. It was about determining if the new scales adequately represent the theoretical conceptualisation of motivational beliefs and states developed during phase one. First, provisional scales were subjected to cognitive interviews (e.g., Willis, 2005) to test whether or not Rwandan students understood the motivational items as intended. Second, exploratory factor analyses (EFA) were employed (Brown, 2015; DeVellis, 2017; Flora & Flake, 2017; McCoach et al., 2014). Factor analysis is used to parsimoniously summarise covariances between observed variables such as self-report items through a lower-dimensional set of latent variables (i.e., factors). EFA in this respect is a procedure that determines the factor structure in a predominantly data-driven way (Fabrigar & Wegener, 2012; Floyd & Widaman, 1995; Preacher & MacCallum, 2003; Reise et al., 2000; Watkins, 2018). Items are expected to load only on the factor the scale they belong to is supposed to represent empirically. Items that exhibit poor associations with that factor or substantial correlations with more than just one factor (i.e., cross-loadings) are revised or removed (L. A. Clark & Watson, 2019; McCoach et al., 2014; Simms, 2008). Within the main phase, the emerging scales were then used to examine the external nomological network of achievement motivation.

The three phases of the present DPhil project can also be described from a construct-validation point of view (Benson, 1998; L. A. Clark & Watson, 2019; Flake et al., 2017; Loevinger, 1957; Simms, 2008). Construct validity refers to the extent to which evidence suggests that the scores collected through a scale reflects the target construct this scale is supposed to operationally represent (Cronbach & Meehl, 1955; Messick, 1995). Accordingly, the first project phase as well as the scale-development stage of the second phase concerned the “substantive component” (Benson, 1998, p. 11) of the instruments on achievement motivation within the Rwandan context. During both steps, the conceptualisation and then empirical operationalisation of the local manifestation of achievement motivation were determined (Benson, 1998; Gehlbach & Brinkworth, 2011; Simms, 2008). Instead of importuning scales from the Western-borne literature, the motivational scales were developed based on actual data from Rwanda, which should have contributed to their cultural validity (F. M. Cheung et al., 2011). The data were directly provided by Rwandan secondary students, the target population of the scales, which should have contributed to the scales’ content validity (Haynes et al., 1995). Content validity is generally assessed through an independent review by experts who are familiar with both the construct content as well as the target group to be studied (Haynes et al., 1995; McKenzie et al., 1999; Messick, 1995). Unfortunately, it was not possible to identify academic experts that are both familiar with achievement motivation and have sufficient command of Kinyarwanda. Instead, cognitive interviews involving provisional scales were performed to ensure that target group members comprehended and interpreted motivational scales in the way intended (“Response Process Validity”; AERA, APA, 2014, p. 15). The scale-testing stage of the second project phase concerned the “structural component” (Benson, 1998, p. 11) of the motivational scales. At this point, the internal patterns associated with the new motivation scales were examined (Benson, 1998; Flake et al., 2017; Simms, 2008). The refined scales emerging from the series factor analyses should therefore reflect key aspects of the hypothesised theoretical conceptualisation of achievement motivation in Rwanda, which implies their structural validity. The third project phase concerned the “external component” (Benson, 1998, p. 11) of the emerging scales. As discussed earlier, it is about locating the scales into the hypothesised inward- and outward-facing external part of the nomological network of the concept the scales are supposed to represent (Cronbach & Meehl, 1955). This step is important because it is only through the empirical relations with other educational variables that the empirical operationalisation of a concept obtains “meaning” (Benson, 1998, p. 11; Crocker & Algina, 2006). The third (i.e., main) phase of the present DPhil project, which was mainly about the external nomological network of achievement motivation within the Rwandan context, is outlined next.

4.2. The project’s main phase: procedure, participants, and classroom observations

The present project's third (i.e., main) phase addressed the second to fifth sets of research questions. It thus concerned especially the level-specific external aspects of the hypothesised nomological network of achievement motivation. During the main phase, relations with external variables were studied not only within the different levels of specificity but also across different levels of specificity. Examples for the former include the hypothesised relations between prior academic achievement and subsequent self-reported motivational beliefs (second set of research questions) as well as between classroom activities and behavioural engagement (research question 5.2). Examples for the latter involve the correlations between self-reported school subject-specific motivational beliefs and aggregated lesson-specific motivational states (research question 3.2) as well as between lesson-specific motivational states and situation-specific classroom activities (research 4.2). It was especially for the cross-level relations that the research design of the main phase covered different levels of specificity as well. It combined not only primary (e.g., motivational self-reports) with secondary (e.g., school grades) data collection but also cross-sectional (i.e., self-reported school subject-specific motivational beliefs) with an intensive longitudinal data-collection framework. Cross-sectional methods imply the collection of data at one specific point in time. Thus, it allows gauging inter- but not intrapersonal differences within variables of interest (e.g., X. Wang & Cheng, 2020). Given that motivational beliefs are assumed to be stable across time, a cross-sectional approach to gauging them was deemed adequate. By contrast, lesson-specific motivation and behavioural engagement observed were hypothesised to be the outcomes of psychological processes each manifesting as a "sequence of states" (Schmitz, 2006, p. 436). To capture the temporal variability within states, lesson-specific motivation and behavioural engagement were measured through an intensive-longitudinal design. Intensive longitudinal studies imply high-frequency in-vivo measurements of focal constructs in close temporal proximity to the events of interest (e.g., Bolger & Laurenceau, 2013; Mehl & Conner, 2012; Walls & Schafer, 2006). By measuring focal constructs in high frequency within short timeframes, researchers can estimate their susceptibility to situational influences. By measuring focal constructs in close proximity to the events of interest reduces recall bias amongst respondents (Bolger & Laurenceau, 2013; Shiffman et al., 2008), something that is especially important of data collection relying on self-reports. Since they often measure constructs in vivo, intrapersonal studies are particularly generalisable to the real lives of respondents (i.e., ecological validity). Furthermore, modelling motivational traits requires intensive longitudinal data-collection designs in order to be able to estimate true interpersonal differences within motivational states that are free form situational influences (e.g., Geiser, 2020). Data collection targeted chemistry and mathematics, two quantitative school subjects, as well as English and Kinyarwanda, two language-based school subjects. Whereas English is the medium of instruction at Rwandan secondary schools, Kinyarwanda is the national language. To control for ability grouping between schools, the main study included not only basic-education schools but also public boarding schools (Williams, 2015). The overall sequencing of the cross-sectional and intensive-longitudinal data collection is outlined next.

Figure 6: The sequencing of data collection during the project main phase



4.2.1. The procedure

The project's main phase was carried out during the first six weeks of teaching of the first school term of Rwanda's 2020 school year. As highlighted in figure 6 above, the cross-sectional data collection preceded the intrapersonal stage (for a similar example, see Ahmed, Minnaert, et al., 2010; Götz et al., 2013). During the cross-sectional phase, questionnaires on motivational beliefs in chemistry, English, Kinyarwanda, and mathematics were administrated. The took place around the time schools completed student registration and started teaching. In Rwanda, however, the time it takes to complete student registration varies across school grades. Across schools, the registration of secondary-1 (S1) students takes around one week longer than the registration of secondary-2 (S2) and secondary-3 (S3) students. To account for the different term starting points, S1 students completed the survey a week later than S2 and S3 students. Due to a glitch, four of the cross-sectional survey questions spoke about the wrong subject (e.g., Kinyarwanda instead of mathematics). Those questions were administered to classes affected by it again the next working day. On data-collection days, questionnaire administration was managed by four survey teams (one for each school), each consisting of two enumerators. Questionnaires were only handed out to students in their classrooms outside study time (e.g., lessons cancelled) in the absence of teachers. After greeting students, enumerators read out questionnaire instructions to students in Kinyarwanda. The purpose was to explain the different sections of the questionnaire, the logic behind the answer options, and to highlight again students' rights during the research process. It took around 21 minutes on average to complete questionnaire administration in class.

The intrapersonal phase targeted English, Kinyarwanda, and mathematics only. Chemistry was dropped to increase the number of lessons that could have been effectively reached per subject given field staff available. Classroom observations were carried out by a team of seven raters (five females). They were supported by a team of four research assistants (two females) who managed the administration of the state questionnaires in class. They were instructed to make sure that teachers did not look at the questionnaires and especially the answers students provided. Teachers that wanted to look at a questionnaire or answers provide by students were "distracted" by research assistants amongst others through engaging them in a conversation. On average, it took around five minutes per class. The procedures for classroom observations are outlined in sub-section 4.2.4. The intrapersonal phase commenced the week after the cross-sectional survey was completed. It lasted for 13 school days (see figure 6). Intrapersonal data collection was interrupted by a public holiday where all schools were closed. Moreover, on one day during the second week all field work was suspended because the printer could not provide all data-collection forms on time. As per the lesson schedules provided by schools, 531 English, Kinyarwanda, and maths lessons were supposed to take place across the three schools during those 13 days. Given field staff available, it was not possible to cover them all during data collection. Instead, 436 lessons were targeted (82.1 per cent), having been randomly selected in a way that ensured a somewhat equal spread across schools, target classes, and subjects. In 58 out of 436 lessons (13.3 per cent) the teacher did not come to class. The teacher absenteeism observed thus certainly reduced the available data significantly since less state questionnaires could be collected, and classroom observations could be carried out. Teacher absenteeism is not unusual within the African context (Chaudhury et al., 2006). Hence, the scope of teacher absenteeism observed during the intrapersonal phase may suggest that the data collection associated with this DPhil project may not have resulted in (large) research response effects amongst teachers. Despite the presence of field teams in schools, teachers continued to be absent from class. This in turn may suggest that the DPhil project did not create a form of "artificial" environment during data collection and thus student reports may indeed be indicative of their usual motivational experiences (i.e., ecological validity). In nine of the targeted lessons (0.2 per cent), subjects different than the one listed in the lesson schedules took place. In five of the targeted lessons (0.1 per cent), other school events prevented lessons from taking place. In 33 cases, logistical problems prevented any data collection. In total, data were collected in 331 lessons (75.9 per cent) took place. In 321 out of those 331 lessons (95.5 per cent), classroom observations were carried out. The lesson-specific questionnaires were administered in 320 lessons (95.2 per cent). In 312 out of the 331 lessons, both types of data were collected (94.3 per cent). In the next section, some general comments on the research support staff employed during this DPhil project are provided.

4.2.2. The research support staff

Local research support staff was employed to support the different activities associated with the DPhil project. Enumerators were employed to administer questionnaires (e.g., see the research outlined in manuscript two in section 5.3). Data entry operators were used to process the data collected through paper questionnaires electronically. Raters were employed to carry out classroom observations (e.g., see the research outlined in manuscript five in section 5.6). In addition, a senior research assistant was employed to help coordinate research logistics, supervise support staff involved in data collection, manage communication with schools as well as local authorities, and to help analyse and interpret research data (i.e., see the research outlined in manuscript five in section 5.6) through the different phases of the research project.[9]

All support staff employed were Rwandans living in Kigali or Bugesera, the two districts where this DPhil project was implemented. They were between 20 to 30 years of age. Almost all had completed university degrees at one of the local universities. Being locals, they were all familiar with public transport in Kigali and Bugesera, which all of them to move from one school to another, if necessary unsupervised. For example, student raters were expected to travel between schools during the intrapersonal phase of the project (see manuscripts four and five) depending on their daily work schedule. Being locals that all attended schools in Rwanda they were also familiar with the school context the DPhil project was operating in. This in turn allowed them to behave appropriately within schools and, if required, to interact with students, parents, and school staff in locally appropriate ways.

Support staff stayed with the project until completion of respective activities. Only the senior research assistant was replaced halfway. The first senior research assistant stayed until the completion of study one of the main phase (see manuscript one). She was a young

[9] Halfway through the completion of the DPhil project, the senior research assistant was replaced. Work started off with a female research assistant, who moved on to attend a master's programme at an international university within Rwanda. She was replaced by a male research assistant who was comparable in terms of English language skills and prior research training.

Rwandan woman with prior experience in qualitative research methods. After completion of study one, she started a master's programme at an international university in Rwanda. For the remainder of the fieldwork, the project was supported by a male senior research assistant who was about to finish a university degree in clinical psychology at University of Rwanda. Hence, he already completed his training in qualitative and quantitative research methods. Both senior research assistants were fluent not only in Kinyarwanda but also English. This in turn allowed them to help communicating with all other project staff as well as external stakeholders such as teachers, parents, students, and local authorities.

Both senior assistants were trained up on the motivational theory behind the design of the DPhil project. This was particularly true for the first senior research assistant. Her training focussed especially on the different facets of expectancy value theory. Within the training, not only the theoretical underpinnings of the expectancy value model were discussed. Another focus of the training was about the way the different expectancy and value facets may operationalize within school contexts. For that purpose, different qualitative studies on psychological phenomena around achievement motivation were discussed. Given her prior experience in qualitative research and her own first-hand familiarity with schooling within the Rwandan context, the first senior research assistant was therefore best prepared to support not only data collection but also data analysis reported in manuscript one. The second senior research assistant was subjected to a very similar training programme, which in turn allowed him to support the analysis of the cognitive interviews carried out during the instrument design phase (see section 5.2).

The classroom raters were involved in finalization of the observation protocol. In this respect, their experience with Rwandan schooling and the way Rwandan students and teachers behave were invaluable to the refinement of the observation protocols including the coding schemes for both student engagement and classroom activities. Upon the development of the observation protocol, the classroom raters were trained up on the observation procedures and coding schemes. As explained in annex five, the training was extensive consisting of several weeks. This in turn allowed standardizing the way the protocol was implemented across the different team members. Within regular intervals through the training, the different coding categories were discussed with the raters and revised based on their feedback. Thus, the training interactions with the local classroom raters helped to further ground the observation protocols within ground-level realities of Rwandan classrooms.

4.2.3. The study participants

The main phase was implemented across four schools from two different districts of Rwanda. The first district (i.e., district A) is within the province of Kigali and thus quite urban in comparison to the second district (i.e., district B), which is located outside Kigali and thus more rural in its geography (NISR, 2016). For each district, one boarding and one basic-education school were selected from the districts' school lists. The school choices constituted a convenience sample in that institutions were selected that are somewhat in close proximity to each other. Otherwise, coordination of survey teams across locations would have been difficult. As is common, both boarding schools only offer secondary education. The boarding school within district A (i.e., school A1) consists of 10 lower-secondary classes and 503 students in total. The second institution in district A (i.e., school A2) is a day school with 495 lower-secondary students across seven classes. The boarding school in district B (i.e., school B1) was the smallest institution with only 259 lower-secondary students across six classes. School B2, a day school, had 11 lower-secondary classes that provided for 554 students. To recruit students, all target classes were repeatedly visited during the schools' student-registration weeks prior to data collection. On those visits, the purpose of the study was introduced, the data-collection procedure was outlined, and rights of participants explained. Participation was voluntary and was not tied to any form of benefit. As a token of appreciation for partaking in the research, all participants were eligible to join a tombola game upon completion of fieldwork. Three winners were randomly selected per class, each winning a pocket calculator worth approximately £0.75. In total, 1638 students were recruited. At school A1, 481 students were accepted into the study (95 per cent) and another 450 at school A2 (91 per cent). At school B1, it was 233 students (90 per cent) and another 476 students at school B2 (89 per cent).

As table 1 highlights, however, participation rates varied across the different data collection exercises. A key factor in this respect was that school A1 decided against continuation of the research at their school upon completion of the first measurement of motivational beliefs.[10] Moreover, both S1 classes at school A2 were removed from the intrapersonal study phase. Student numbers in both classes were around 100, which was deemed too much to ensure any reliable classroom observations. Across all four schools, 1384 students (90 per cent) completed the first measurement of motivational beliefs (i.e., the baseline). Across the remaining 22 classes, 1310 students (79.9 per cent) submitted at least one lesson-specific questionnaire. The average number of lesson-specific questionnaires submitted was 9.8 ($SD = 4.74$; $range = 1 - 24$). Classroom observations targeted the same 10 (focal) students per class, which were randomly selected from the class register before the beginning of the intrapersonal phase. In the case of an S3 class of school A2, one of the focal students was not found in class. Initially, it was not clear whether or not the student had dropped out of school. Thus, the student was not replaced. In total, 219 focal students were subject to at least one set of classroom observations. The average number of lessons during which students were observed is 7.4 ($SD = 4.45$; $range = 1 - 21$). Table 1 on the next page provides a breakdown of student participation across all data collection exercises.

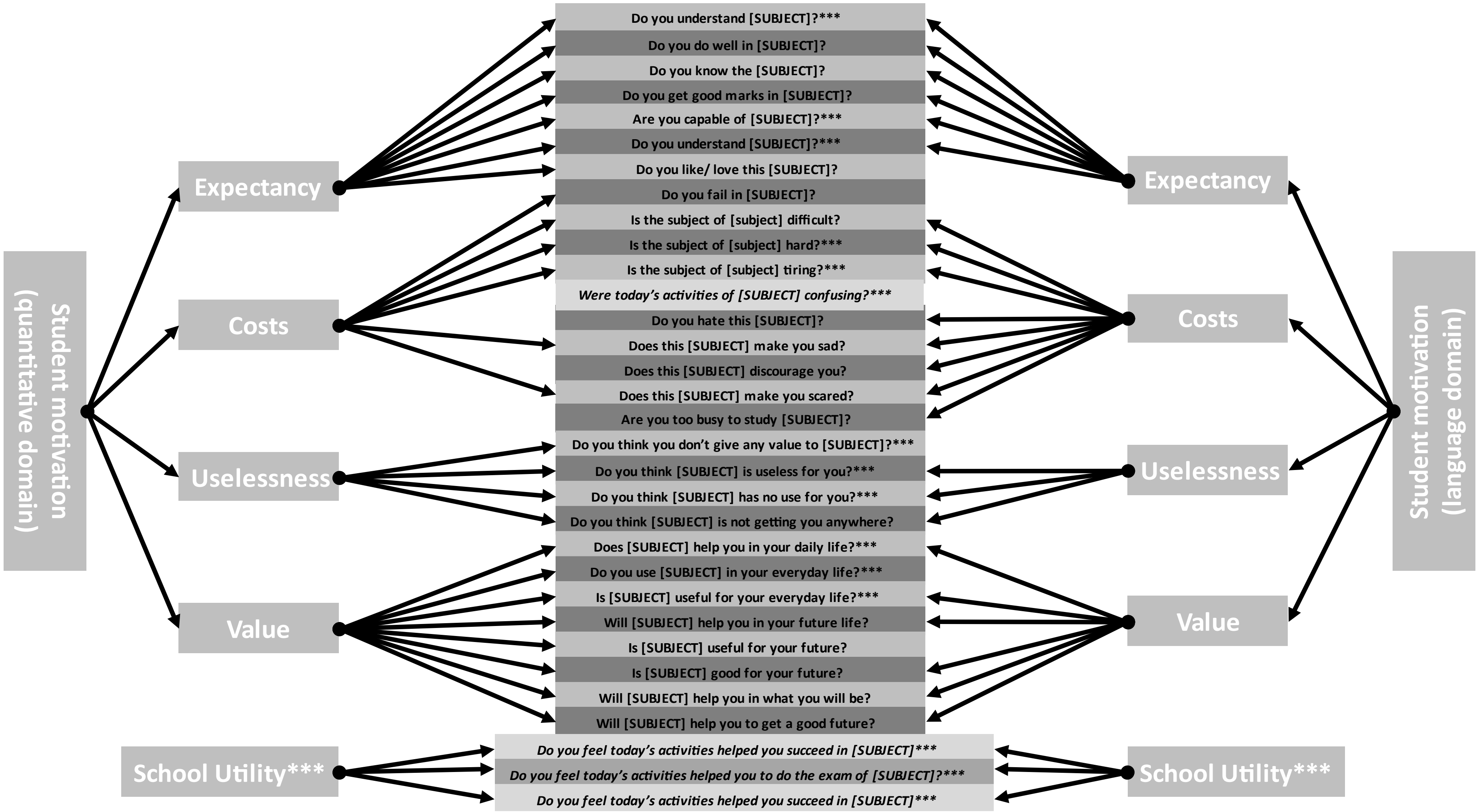
[10] According to the school administration, classroom attendance of student observers especially would have disrupted classroom-based teaching, something they wanted to avoid. Schools dropping out of research is not that uncommon. For another example see Li and Learner (2013).

Table 1: Full sample composition (project main phase)

Participants	A1 (Boarding School)			A2 (Day School)			B1 (Day School)			B2 (Boarding School)			Total
	S1	S2	S1	S1	S3	S3	S1	S3	S3	S1	S3	S3	
# of classes	3	4	3	2	2	3	4	4	3	2	2	2	34
# of students	156	205	142	191	143	161	230	174	131	81	109	69	1792
# of participants (cross-sectional)	148 (43%)	179 (42%)	120 (50%)	137 (48%)	107 (50%)	112 (55%)	161 (55%)	121 (63%)	100 (55%)	69 (33%)	77 (51%)	53 (49%)	1384 (50%)
# of participants (repeated)[1]	NA	NA	NA	187 (45%)	140 (50%)	161 (55%)	232 (50%)	181 (54%)	155 (52%)	82 (33%)	104 (47%)	68 (54%)	1310 (49%)
# of focal participants[2]	NA	NA	NA	NA	20 (50%)	29 (49%)	40 (48%)	40 (55%)	30 (54%)	20 (41%)	20 (52%)	20 (53%)	219 (50%)
# of participants for whom school grades available	9 (44%)	124 (44%)	120 (50%)	8 (12%)	100 (49%)	108 (55%)	53 (62%)	100 (61%)	93 (57%)	1 (100%)	53 (53%)	53 (49%)	822 (52%)

Note. Number in brackets refer to the percentage of females. [1)] These participants are students that have submitted a questionnaire on motivational states at least once during the intrapersonal data-collection phase; [2)] These participants are those students who have been randomly selected to be subjects of classroom-based observations and that were actually found in class at least once during the intrapersonal data-collection phase.

Figure 7: Motivational scales across levels of specificity and school subjects (main phase)



Note. 'Do you understand [subject]' is included twice. This is due to translation of the Kinyarwanda words 'gusobanukirwa' (i.e., to understand/ to be explained to) and 'kumva' (i.e., to understand/ to listen/ to hear).

4.2.4. The motivational scales

Section 5.2 below outlines the series of steps taken to derive both the lesson-specific as well as school subject-specific scales. Since the scale development and testing phase were data driven (i.e., inductive) and carried out for each level of specificity, solutions were derived that partly varied across subjects as well as level of measurement. Figure 7 above compares scales across subjects and level of specificity.[11] The dimensions of school subject-specific self-reported beliefs were measured through scales on expectancy, daily utility, and costs. They slightly varied in terms of the length and item composition between the quantitative (i.e., chemistry and mathematics) and language domain (i.e., English and Kinyarwanda). This was mainly due to the fact that the scales on school subject-specific self-reported beliefs were developed separately for each domain. This was not done in case of the lesson-specific scales. Thus, they did not vary across school subjects. The dimensions of lesson-specific achievement motivation were assessed through four scales, each three items long, namely competence, daily utility, school utility, and effort perceptions. In case of lesson-specific self-perceptions, the scale was not called expectancy perceptions but competence self-evaluations. The former perceptions are predominantly future bound. In the case at hand, self-perceptions were measured upon completion of a lesson and were thus retrospective. It is for this reason that they were called competence self-evaluations and not expectancy perceptions (Perez et al., 2014). The lesson-specific questionnaire was complemented by an item on perceived novelty. Novel learning experiences have been associated with engagement (Durik & Harackiewicz, 2007; Renninger & Bachrach, 2015).

[11] As figure 7 highlights, competence beliefs are operationalized through two items that are translated into English through 'Do you understand [subject]'. Within Kinyarwanda, these items are different in the sense that they use different verbs. The first is making use of the verb 'gusobanukirwa' (i.e., to understand/ to be explained to) whereas the second makes use of the verb 'kumva' (i.e., to understand/ to listen/ to hear). Thus, both items do indeed capture different aspects of the Kinyarwanda semantics around competence beliefs. Also see annex 3 for the wording of the scales in the original.

All items relied on the three-point response format. This is not too common within research of psychological constructs. Psychometric tests generally suggest that the reliability and validity of scales on psychological constructs is maximized when the number of response options is between four and seven (e.g., DeVellis, 2017; McCoach et al., 2014). Response formats with respective numbers of response categories tend to provide respondents with the opportunity to sufficiently distinguish between the different nuances within their psychological experiences without being overly complex. Against this background, a three-answer response format could be seen as less favourable from a scale-development perspective. However, respective psychometric tests were performed based on the ability to construct multi-response categories that made sense to respondents in the first place. In the context of this DPhil project, systems with four- and five-response categories were tested with Rwandan secondary students. For example, the four-response format consisted of 'agree', 'somewhat agree', 'somewhat disagree', and 'disagree'. In general, respondents interviewed during the cognitive testing considered the two answer options on each side (e.g., 'agree' and 'somewhat agree') to be very similar. Furthermore, they often suggested to provide a simple yes/no answer format rather than four or more answer options. To provide respondents with the possibility to express higher level of granularity than a simple yes/no, it was therefore decided to include a three-options response format. Interestingly, student answers often tend to be binary, nonetheless. As figure 12 on page 176 highlights, students tend to choose 'sometimes' and 'often' most of the time. By contrast, 'never' was hardly chosen. Future research needs to further explore the cognitive processes around these phenomena amongst Rwandan students. Both the school subject-specific and the lesson-specific questionnaire started with questions on personal information (e.g., name, age, gender, class), which helps to track respondents across the different measurement occasions.

Table 2: Average polychoric correlations for all lesson-specific scales (main study)

Scale	English	Kinyarwanda	Mathematics	Average
average within-level polychoric correlations (lesson-specific self-reports)				
Competence	0.59	0.54	0.62	0.58
Daily utility	0.52	0.55	0.58	0.55
School utility	0.55	0.44	0.54	0.51
Cost	0.52	0.54	0.52	0.53
Average	0.55	0.52	0.57	
average between-level polychoric correlations (person-specific self-reports)				
Competence	0.98	0.96	0.97	0.97
Daily utility	0.96	0.98	0.98	0.97
School utility	0.89	0.93	0.91	0.91
Cost	0.97	0.98	0.95	0.97
Average	0.95	0.96	0.95	

Note. Polychoric correlations were estimated using a multilevel framework. Level one represented the lesson-specific within-person level and was nested within level two which constituted the between-person level.

A comment is warranted. Not all data that was collected through all four lesson-specific motivational scales were used in the analyses presented in manuscripts three to five. The application of the GI/E model to average motivational states only concerned the competence and cost scales. Models including daily utility and school utility did not converge. The analyses in manuscript four and five only included the competence, daily-utility, and effort (cost) scale. Again, models including school utility did not converge. Since school utility was not used in any of the five manuscripts, table 2 and 3 is included above. Table 2 presents the average polychoric correlations across items within each scale and domain as a measure of reliability.[12] Amongst lesson-specific scales, the average correlation was .54 at the at the within level and .95 at the between level. Thus, scales generally exhibited reliability. However, school utility tended to exhibit slightly lower scale reliability compared to the other three scales. For example, whereas all other scales showed an average correlation of .97 at the between level, it was only .91 for school utility. Table 3 presents

[12] Within the literature, no consensus has yet been reached on how to compute reliabilities of scales involving categorical items with more than two items. It is for this reason that we relied on polychoric correlations to obtain at least a crude measure of reliability rather than any of the other, commonly used but necessarily more reliable measures.

Factor means, variances and ICCs for all four lesson-specific motivational scales including school utility. Compared to the other scales, school utility exhibited rather low within and between-lesson variances. This in turn suggest that responses to the school-utility fared more similar between lessons as well as between students as they did for the other three scales. School utility thus exhibited slightly worse psychometric properties than the other three scales. This in turn may explained why structural models including school utility did generally not converge.

Table 3: Factor means, variances and ICCs for all lesson-specific scales (main study)

Scale	Within-level variance	Between-level mean	Between-level variance	ICC
		English		
Comp	**2.57 [1.99; 3.52]**	**3.09 [1.28; 4.82]**	**4.13 [3.12; 5.69]**	.62
Daily	**1.42 [1.02; 2.04]**	**3.35 [1.78; 4.71]**	**3.30 [2.34; 4.86]**	.70
School	**.92 [.70; 1.23]**	**2.50 [1.03; 3.65]**	**1.96 [1.47; 2.69]**	.68
Cost	**1.62 [1.32; 2.04]**	1.92 [-.57; 4.30]	**2.73 [2.19; 3.50]**	.63
		Kinyarwanda		
Comp	**2.44 [1.56; 4.08]**	**3.27 [.57; 6.14]**	**3.84 [2.36; 6.29]**	.61
Daily	**1.98 [1.37; 3.23]**	**2.74 [.20; 4.43]**	**4.85 [3.28; 7.94]**	.71
School	**.73 [.44; 1.24]**	**2.90 [1.46; 4.29]**	**1.72 [1.03; 2.97]**	.70
Cost	**2.79 [1.90; 4.07]**	**2.15 [.12; 5.06]**	**5.43 [3.71; 8.09]**	.66
		Mathematics		
Comp	**2.46 [2.03; 3.10]**	2.02 [-.54; 3.44]	**2.82 [2.25; 3.64]**	.53
Daily	**1.55 [1.27; 1.93]**	**2.53 [1.36; 4.47]**	**4.21 [3.35; 5.33]**	.73
School	**.83 [.67; 1.01]**	**2.09 [.83; 3.14]**	**1.66 [1.32; 2.12]**	.67
Cost	**1.73 [1.42; 2.14]**	1.39 [-.37; 2.73]	**1.96 [1.59; 2.46]**	.53

Note. ‘Comp’ = competence states; ‘Daily’ = daily-utility states; ‘School’ = school-utility states; ‘Cost’= cost states. ‘ICC’ = intracluster correlation. Sample size = 868 (see also manuscript four); Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits were general good: PPP ranged from 0.337 to 0.513. f (95% CI) always included zero. Coefficients in bold are significantly different from zero.

4.2.5. School grades

In the context of this DPhil project, academic achievement was of interest especially in the context of testing the GI/E model. Thus, prior academic achievement, and its effects on subsequent motivational beliefs, were therefore in the centre of the study. Prior academic achievement was operationalised through final school grades from the previous school year (e.g., Gaspard et al., 2018). Rwandan students are used to school grades as measures of academic progress. School grades may therefore be more predictive of subsequent motivational beliefs than, for example, one-time external learning assessments (Marsh, 1987; Möller et al., 2009). In Rwanda, end-of-year school marks are generally presented as per centages with 100 per cent indicating perfect achievement. End-of-year marks represent averages students attain across the three school terms that make up the Rwandan school year. First- and second-term grades are generally determined through school exams and tests organized locally. School exams within the third term are prepared by the national level and uniformly administered across schools in the country (Nzabalirwa, 2014). In Rwanda, end-of-year school marks can therefore be considered somewhat comparable across schools. A major caveat of school grades is that they may not necessarily be indicative of actual academic progress. Sandefur (2018) observed that African countries generally tend to lack behind OECD countries in terms of academic achievement. This may not be reflected within school grades given out to African students. Thus, school grades in counties such as Rwanda may constitute a benchmark of academic achievement that far too often is not indicative of academic progress one would expect students within a certain age bracket to have attained.

4.2.6. The classroom observation tool

Both situation-specific classroom activities as well as the level of behavioural engagement exhibited during these activities were gauged through classroom-based observations by external raters (Godwin et al., 2013; Godwin & Fisher, 2011; Rosário et al., 2016; Seidman et al., 2018). The protocol that guided the in-situ observations of both classroom activities and student engagement was developed through a multi-staged approach consisting of a literature review to derive a provisional protocol, extensive field tests of the protocols, and incorporation of feedback by external raters. Classroom ratings of engagement and activities were based on systematic observations (Salvia et al., 2007; Sanders et al., 2010; Wragg, 2012) carried out by external raters located in class (Godwin et al., 2016, for studies wth a similar observation design see 2013). This approach is considered adequate if the focus is on frequency rather than the quality of "specific behaviours" (Sanders et al., 2010, p. 425). To guide ratings, a coding frame for both student engagement (see table 4) as well as classroom activities (see table 5) were developed (see annex 5). The operational definitions of the categories within each of the two coding frames were derived from the literature, field-tested, and adjusted until each coding scheme was found to adequately capture the typical range of lesson activities and students' overt behaviours in Rwandan classrooms.

Table 4: The engagement coding scheme (observation protocol)

Level	Description and Examples
Verbally on-task	The student is not only paying attention to the classroom activity by having their body and eyes oriented towards a relevant learning object but is also making (or wanting to make) verbal contributions to a lesson in the form of answers to questions, clarifying questions, suggestions, etc. A typical example is a student who wants to answer a question asked by the teacher and therefore has their hand raised while looking at the teacher. As is common in Rwanda, the student may also be saying "teacher me!".
On-task	The student is paying attention to the classroom activity by orienting their body and directing their eyes towards a relevant learning object (e.g., teacher talking, student talking, or blackboard with notes). A student may be active in that they are copying notes or working on a task. However, the student is not verbally active in that they are asking any questions or wanting to answer a question.
Off-task	The student is not paying attention and their eyes are thus directed towards anything but the learning source. Their body may also be oriented away from the classroom activity. However, even though the student is not paying attention, they are not disturbing the lesson either. Typical examples include a student who is staring out of the window or a student whose head is resting on the bench with the eyes closed.
Off-task disruptive	The student is disrupting the lesson through chatting, laughing, making noise, and/or any unnecessary movements. In addition, their eyes and posture may also be oriented away from the classroom activity. A typical example includes a student who is talking to a neighbour about something that is not related to the lesson.
Waiting	No lesson activity is currently ongoing and thus the student cannot be expected to be engaged with any specific task. It does not imply situations in which

	teachers are absent while students have learning tasks to complete. A typical example is a class whose teacher is outside the classroom to talk with a colleague leaving students idle. Students are therefore chatting, sleeping, or are moving around in class.
Absent	Even though attending school that day, the student is not in class. A typical example is a student who was called out of class by the school secretary to discuss school administrative issues.
Cannot be observed	The student cannot be observed because the view of the student's face and especially eyes is blocked. A typical example is a student who sits in a bench behind a group of students whose heads and bodies are entirely blocking the view of the student.

Note. Based on Hospel et al. (2016), Ocumpaugh et al. (2015), and Shapiro (2011).

The coding frame on engagement centred around passive (on-task) and verbal behavioural engagement (verbally on-task; Hospel et al., 2016; Ocumpaugh et al., 2015; Shapiro, 2011). Verbally on-task was considered to be nested within on-task behaviour in the sense that in order to be verbally engaged students also had to be on-task (e.g., Shapiro, 2004). In addition, the categories of off-task, and off-task disruptive were included (Shapiro, 2004).[13] On- or off-task was determined especially through the student's gaze (i.e., visual attention) and head orientation. A person's gaze and head orientation have been found to be strong predictors of attention (e.g., Stiefelhagen, 2002), something that seems to vary across cultures only in subtle ways. For example, Chua et al. (2005) found American and Chinese participants to exhibit different eye movement patterns when exploring cluttered photos. The former tended to focus more on the focal objects depicted in the study photos whereas Chinese students also paid attention to background details. Besides this, both groups cognitively engaged with the photos by having their eyes directed towards them. Hence, culture may influence where exactly on an object a person's eyes may wander. However, culture may not influence whether or not the person's eyes are directed towards that object as

[13] One could argue that students could also be on-task disruptive. For example, students shouting the solution of a task into the classroom. Even though such on-task behaviour may indeed be disruptive, it would have been coded as verbally engaged within the current observation scheme. Within Rwandan classrooms, norms around hand raising often did not seem to be well enforced. As a result, students were frequently found to speak out in class right away while not necessarily intending to disrupt the lesson flow.

a whole as long as it constitutes the centre of their attention. It may therefore be universally true that what is engaging is “what catches your eye” (Reeve, 2012, p. 149). Whereas engagement often comes with eye fixation on the centre of attention, catching someone’s eyes does not always lead to attention and engagement. Mind-wandering, too, has been associated with having one’s eyes rested on an object (Faber et al., 2020). Despite this caveat, eye gaze is generally considered a “reasonable (albeit imperfect)” (Godwin et al., 2013, p. 2) measure of attention (also see Henderson & Ferreira, 2004; Just & Carpenter, 1976). Within the present DPhil project, a student was rated on-task when they were looking, not staring, towards the teacher, the blackboard, their own notebook, or any peer doing something of relevance to the lesson (Godwin et al., 2016). Otherwise, the student would have been coded off-task. Being verbally engaged implied not only being on-task but also making (or wanting to make) a verbal contribution to the lesson (Reeve & Tseng, 2011). During observations, raters determined the nature of a student’s behavioural state but not its duration, intensity, or quality (Alberto & Troutman, 2013). In situations where students were left idle (e.g., teacher has left the class) they were coded as ‘waiting’. The coding frame also allowed to indicate when students did not face any specific activity to perform (i.e., waiting), being absent, or when face and eyes were not visible (i.e., not observable).

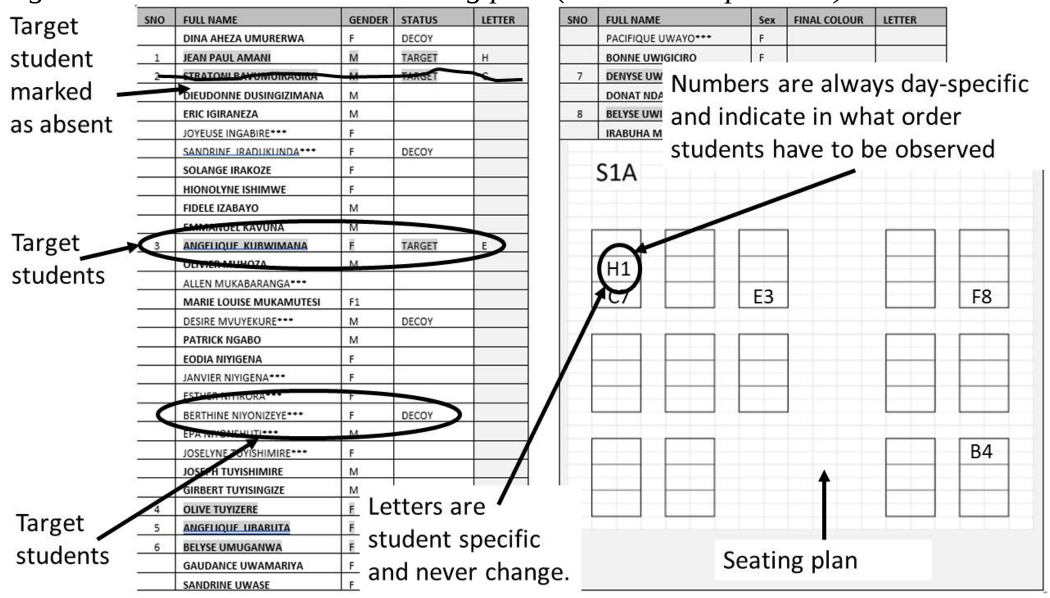
The level of engagement exhibited also depends on the nature of activities they are expected to perform. To control for the lesson context, student ratings were coupled with ratings of the prevailing classroom activity during the observation interval (Shapiro, 2004). To categorise classroom activities, the World Bank’s (2015) revised version of Stallings’ (e.g., 1977) Classroom Snapshot Observation System was adjusted to the Rwandan context. Stallings’ system has already been applied to the developing-country contexts (e.g., Abadzi, 2007; Adelman et al., 2015; Venäläinen, 2008). As table 5 on the next page highlights, these classroom-based activities can be loosely divided into three blocks. They can be instructional, managerial, or unrelated to any direct teaching objective (World Bank, 2015).

Table 5: The classroom-activity coding scheme (observation protocol)

Category	Learning Events	Description
Instructional activities	Reading aloud	The teacher is, or student(s) are, reading aloud a text from blackboard, notebook or textbook while the rest of the class is expected to follow or read along in chorus.
	Teacher explanations	The teacher is explaining or demonstrating tasks, homework, or any other academic content to the class with the intention of making students understand it. Alternatively, student(s) are giving presentations to the class.
	Group work	All students in the class are divided into groups where they are expected to work on a group assignment. In most cases, students rearrange their seats in accordance with how they have been divided into groups. During group work, the teacher generally goes from group to group to provide support and guidance.
	Student-based demonstrations	Student(s) are (silently) solving an exercise in front of the class that has been written on the blackboard while the teacher and the class watch the task completion. It does not imply students explaining a task.
	Copying teacher's notes	Students are copying notes that the teacher is writing on the blackboard. At times, the teacher may read out passages written on the board, mainly to help students copy notes properly.
	Seat work	All students in class are working on a given task by themselves. In most cases, students are expected to hand in a paper with their answers for the teacher to mark it.
	Copying student's writings	A student is witing notes, normally from a textbook, on the blackboard while all other students in class are expected to copy these notes into their notebooks. In most cases the teacher is not in class.
	Verbal interactions	A student is verbally engaging with the teacher or other students to ask questions, to answer questions, or to discuss academic content while the rest of the class is listening.
Managerial		The teacher is performing the non-instructional activities that ensure students are organised, not disruptive, and on-task. Examples include checking attendance and calling out or punishing disruptive students.
Lesson unrelated	Social interactions	The teacher is interacting with one or several students in a way that is purely social in nature (e.g., exchanging jokes or news) while students have no learning activities to work on.
	Teacher disengagement	The teacher is talking to a visitor (e.g., parent or colleague) or is on the phone while students have no learning activities to work on.
	Teacher absenteeism	It refers to the teacher having left the class while students have no learning activities to work on.

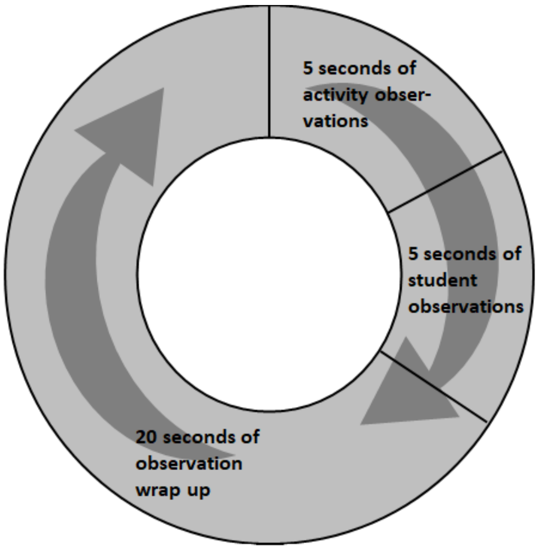
Note. The coding scheme adapted from World Bank (2015). Also see Stallings (1977).

Figure 8: The student list and seating plan (observation protocol)



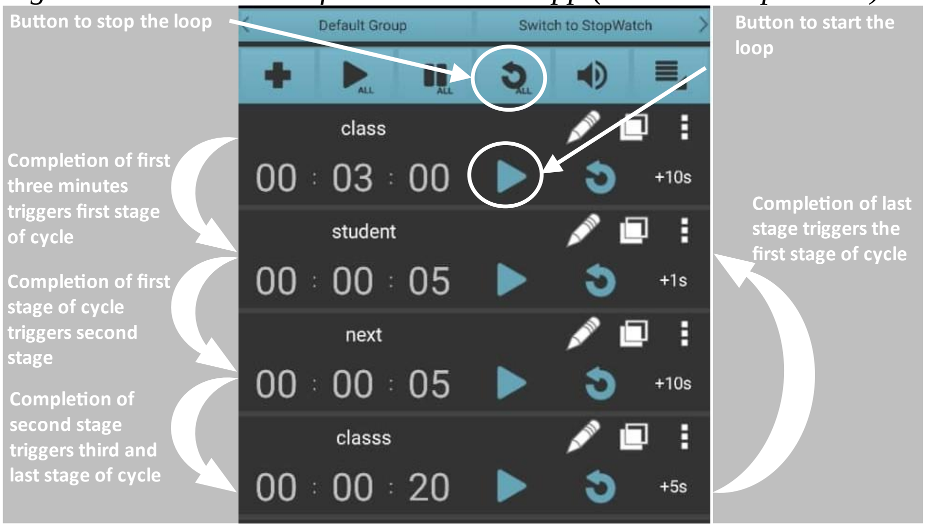
Classroom observations were always carried out by one rater per class. In the minutes between two lessons, the rater entered the class. While waiting for the teacher, the rater located focal students using a class-specific student list. The only feasible way to do so was to read out their names. To minimise any reactivity effects (Walls, 2013), observers pretended to want to check school attendance only. Moreover, the names of the focal students were read out along with the names of 10 other students (i.e., decoy students), whose composition varied across days. The sole purpose of the decoy group was to divert attention from the fact that observations were always targeting the same 10 students per class. Raters were all equipped with an exact seating plan for each class they attended (figure 8). They entered the exact location of focal students in class into the seating plan, which then served as a reminder as to where to find all the class members they have to observe. Focal students not found in class were marked as being absent. In order to obtain a measure of the reliability of this student-tracking approach, a verification step was introduced after the fourth day of data collection. To verify focal students once a lesson was over, research assistants in charge of handing out the lesson-specific questionnaires checked the names of the students that should have been focal students according to the lesson's seating plan. Again, to divert attention, the names of neighbouring students were also enquired. Focal students that were not verified were highlighted in the observation sheet accordingly.

Figure 9: The structure of the classroom-based observation cycle (observation protocol)



The classroom-based observation protocol employed a round-robin design (Ocumpaugh et al., 2015) and organised around a 30-second long three-staged observation cycle (see figure 9). Within a round-robin design, observational data can be collected from a group of participants by rating each in turn (Alberto & Troutman, 2013). To prevent any order effects, the sequence in which to rate students was always randomly determined before the lesson. During the five seconds of the first stage, the rater observed and rated current lesson activities using the classroom-activity coding frame (see table 5). During the five seconds of the second stage, the rater then observed and rated the engagement level of a given focal student using the behavioural coding frame in table 4 while taking into account what behaviours can be expected given prevailing classroom activities (Godwin et al., 2016). Again, to avoid reactivity effects, raters avoided direct eye contact and instead relied on side glances. During both coding stages, the classroom activity and student behaviour that accounted for the biggest proportion of the interval (i.e., predominant-activity sampling; Hutt & Hutt, 1970, p. 69) were chosen. Observation protocols that determine behaviour based on the last second during an interval (i.e., momentary time-sampling approaches) may yield particularly accurate results (e.g., Ferguson et al., 2018) but only if target behaviours are low inference such as sitting or standing in class (Altmann, 1974, p. 259). During the last 20 seconds of a cycle, the observer then entered results of both coding stages into the observation sheet before identifying the next focal student to be observed using the seating plan. To support their classroom observations, each research assistant was equipped with clipboards, the observation sheet, a class-specific seating plan, a smart phone with a multi-timer app, and headsets. The multi-timer app (see figure 10) was programmed in such a way that it gave an audible signal to a research assistant via earpieces whenever a stage within the observation cycle ended. The loop repeated itself until switched off. Within Rwandan public schools, it is common for students to require some time to settle in (e.g., getting notebooks out, etc.) once the teacher started the lesson. Thus, classroom observations always started exactly three minutes after the teacher started the lesson.

Figure 10: Screen shot of the multi-timer app (observation protocol)



A key concern associated with the general design of the observation tool is that students might have altered their behaviours in class because of being watched ("observer reactivity"; Merrell, 2013, p. 115). Despite the use of decoy students, given that their names were read out in every class subject to observations, focal students might have been aware that they were subject to some form of monitoring. However, since every focal student was subject to five-second observations just every couple of minutes, it must have been difficult for them to predict when exactly they were being watched, especially when considering the fact that a researcher's eyes were constantly wandering around the class. Moreover, all research assistants were instructed to never share any research details with classes. Hence, even though the focal student might have sensed they were being watched, they should not have known the actual objectives behind these classroom observations. In addition, reactivity effects have been observed especially in cases where the focus of observations was negative behaviour (Korotitsch & Nelson-Gray, 1999). Hence, if at all (Korotitsch & Nelson-Gray, 1999; Shiffman et al., 2008), any behaviour changes may have materialised especially in the form of focal students being less disruptive to the class. This was deemed acceptable given that disruptive behaviours are not the focus of the study. To support the reliability of observation results, all observers completed an extensive school-based training of more than 15 full days prior to study, which is more than what has been reported for other observation protocols (e.g., Volpe et al., 2005). All observers were also subject to reliability checks throughout the training and study. During reliability checks, classroom-based observations were carried out jointly but separately by the observer to be evaluated as well as this DPhil student. By calculating Cohen's Kappa (Cohen, 1960), the interrater reliability of the results was determined (Hintze, 2005).

4.3. Lower-secondary students: the project's final target group

A final comment is warranted. Within this project, the target eventually shifted to lower-secondary students. There has been no compelling theoretical reason to choose lower secondary over any other school level. However, prior pilots suggested it may be more feasible to carry out data collection on achievement motivation with lower-secondary students than with primary pupils. Primary classes are more crowded, which would have rendered classroom observations challenging. Primary students in Rwanda generally have poor reading skills, which may have rendered the data self-report on their motivational experiences more prone to measurement errors. Finally, it was assumed that social-desirability biases within answers are stronger amongst primary students compared to secondary students (see annex 1).

4.4. General measures to ensure data quality

Quality of research is often discussed in terms of reliability and validity. At the study-specific level, the former has been discussed especially in terms of measures ensuring intersubjectivity in coding (phase one) as well as classroom observations (phase four). To ensure construct validity of all motivational measures, the corresponding scale-development process was organised in such a way that it contained a substantive and a structural phase. To attain content validity of motivational constructs, they were operationalised theoretically and empirically. These were predominantly based on qualitative construct descriptions (i.e., cultural validity) elicited from lower-secondary students in Rwanda (i.e., content validity). Especially the in-depth hands-on analysis of the raw data in terms of the motivational themes as well as language used to express motivation helped this DPhil student to become thoroughly familiar with achievement motivation within the Rwandan context (Berry, 1989). Cognitive interviews further helped to assess the adequacy of emerging scales (i.e., Response Process Validity). The application of EFAs during the structural phase further refined scales and helped to ensure structural validity of all remaining scales. All project phases were implemented in close collaboration with local research assistants. They did not only provide translation services but also provided their views on how to interpret student answers and descriptions from a Rwandan perspective (Bradby, 2002; Walton et al., 2015). This in turn helped to keep cultural bias to a minimum (Mertens, 2010) and to contribute to the face validity of the motivational measures and data-collection procedures (Neuman, 2014). Feedback by research assistants was particularly important for the development of the observation protocol. For that purpose, regular feedback rounds were included at all stages of the development process. As research assistants eventually explained, even from a Rwandan point of view the operationalisations “intuitively [make] sense”. All field data collected were entered into the computer twice to control for data-entry errors during the process.

Moreover, the presence of researchers asking questions about motivation and silently observing school classes might have been an unusual experience for many students (McWilliam et al., 2009; Punch, 2002). This may be particularly true given that these researchers were led by a white foreigner, whom students might have thought, at least initially, to work for a development organisation (Williams, 2015). These misperceptions could have resulted in students wanting to present themselves as particularly motivated both in terms of their questionnaire answers as well as classroom behaviours (e.g., social desirability; Johnson & van de Vijver, 2003). These threads to the ecological validity (Neuman, 2014) of the project data were addressed through the following measures. First, during all research introductions to schools, teachers, and students it was always highlighted that research participation would not lead to any material benefits. Second, especially during classroom-based data collection, all direct student interactions were overseen by local research assistants. This DPhil student mainly stayed in the background in order to avoid any unnecessary attention (Williams, 2015). Third, to prevent any survey reactivity effects (D. T. Campbell & Stanley, 1963), research objectives and other details were discussed neither with students nor with teachers (Rosenthal & Jacobson, 1968). In fact, given the extent of teacher malpractices observed in classrooms during project implementation such as showing up in class late or not at all, one is led to believe that at least teachers may not have substantially altered their teaching practices during the research. Fourth, whenever questionnaires were given out, students were reminded that survey questions were not probing for any "right" answer but only elicited their actual experience with a given lesson or subject. Despite these mitigating measures, especially the unnatural presence of research assistants in classrooms and the social status of "wealth and cosmopolitanism" often attributed to white foreigners in Rwanda (Williams, 2015) may have impacted the overall ecological validity of the project to a certain degree nonetheless.

4.5. Ethics

This DPhil study underwent different review processes all located at different levels. First, the overall research design was reviewed and approved by the Central University Research Ethics Committee (CUREC) at the University of Oxford. CUREC protocols 15 and 25 are applicable to this study. Second, the project design was also reviewed and approved by the Institutional Review Board (IRB) of the College of Medicine and Health Science (CMHS) within the University of Rwanda. Final in-country research approval was granted by Rwanda's National Council of Science and Technology upon further review of the research design. Before schools were contacted, district-level approvals were requested from and granted by those districts covered by the research.

This is an opt-in research project. First, permission was sought from the schools' headmasters. All schools were informed about the scope of data collection and about their right to withdraw from the research at any point. One of the four schools selected for study four decided not to participate in study 4's intrapersonal phase. Next, students were informed about the research during first visits to each of the target classes. On these occasions, consent forms were distributed, all in Kinyarwanda. According to the initial design, research participation required the submission of both a parent consent and a child assent form. The first studies, however, revealed low response rates especially amongst parents. As headmasters informally explained, this might have also been due to parents' low educational status and their lack of exposure to any kind of research. As a matter of fact, nationally representative data suggest that amongst 40–44-year-old Rwandans, which is probably around the age group of parents of most students within lower secondary, more than 90 per cent have completed primary schooling at most. Furthermore, around 29 per cent are considered illiterate (NISR, 2016). It was therefore assumed that low response amongst parents was largely due to a lack of comprehension on the side of parents. In the case of boarding schools, furthermore, it was impossible to even hand out consent forms to most parents in the first place since students do not stay with them. It was therefore decided to have schools act *in loco parentis* in that they provided overall permission to have any of their lower-secondary students be recruited for the research project as long as the corresponding individuals provided assent. In addition, consent forms allowed students to indicate if they wanted the research team to contact their parents about the research. Around 89.9 per cent of students chose this option. In those cases, parents were contacted via phone using numbers provided by students and school. Some parents also preferred to meet the research team in person at school. Around 2.5 per cent of parents opted against their students being part of the research. This alteration of the consent seeking process was discussed with and approved by CUREC. All official interactions especially with district and school personnel as well as with students were only held in Kinyarwanda.

Despite general research approvals, efforts were made to keep any disruptions of classroom-based learning and teaching to a minimum. First, an important aspect of all piloting was to ensure that completion times for crucial aspects of data collection (e.g., lesson-specific questionnaires, student identification during classroom observations) was kept to a minimum. All questionnaires were only given out to students once teachers had indicated that teaching was over. Given these mitigating measures, the potential impact on non-participating students might have been minimal. At times, non-participating students submitted (lesson-specific) questionnaires as well. These forms were accepted but not processed during data entry and analysis.

All collected data have been anonymised and kept confidential according to the Data Protection Act (1998) and in line with the British Psychological Society's (2018) code of ethics. None of the personalised data were shared with anyone outside the research team. All data were entered only by members of the research team. All research team members are committed to confidentiality concerning any details of the research. Once all data were entered and merged, all personalised information was removed from the data set and replaced by unique identification numbers. All data are stored on this DPhil student's encrypted and password-protected laptop, with a back-up on the supervisor's computer. All approval letters, questionnaires, consent forms, and personal information of participants are securely stored and will be discarded five years after the project. All anonymised electronic datasets will be kept indefinitely.

5. The Results

The results obtained through this DPhil project will be presented through five manuscripts that have partly already been submitted to peer-reviewed academic journals. They are presented in sub-section 5.1, and 5.3 to 5.6. Their bibliographical information is provided below. In sub-section 5.2, both the procedures as well as results for developing and testing the scales that were used during the main phase (manuscripts two to five) are described. Scale development and testing of the school subject-specific, self-reported motivational beliefs have been submitted along with manuscript two in the form of supplemental material.

Bulla, D., Uwitonze, A. G., Malmberg, L.-E.,, and Strand, S., (2022a). The cognitive-affective structure of "icyizere" (confidence) and "ishyaka" (motivation) among Rwandan students. A mixed-methods study. *International Journal of Educational Research*. **(Status, 23rd October 2022: major revision to be sent back by 24th October 2022)**.

Bulla, D., Malmberg, L.-E., & Strand, S. (2022b). The Generalized Internal/External Frame of Reference Model among Secondary School Students in Rwanda. *Contemporary Educational Psychology*. (**Status, 23rd October 2022: submitted and currently under review)**

Bulla, D., Malmberg, L.-E., & Strand, S. (2022c). The generalized internal/external frame of reference model with average motivational states. Findings from Rwanda. *Educational Research and Evaluation*. (**Status, 23rd October 2022: submitted and currently under review)**

Bulla, D., Malmberg, L.-E., & Strand, S. (2022d). Does the time allocation across classroom activities predict motivational states? Findings from Rwandan secondary schools. *Learning and Individual Differences*. (**Status, 23rd October 2022: submitted and currently under review)**

Bulla, D., Malmberg, L.-E., & Strand, S. (2022e). Is being on-task predicted by motivational self-reports and classroom activities? Results from an observational study in Rwanda. *Learing and instruction*. (**Status, 23rd October 2022: submitted and currently under review)**

The contributions of each author (CRediT) to the manuscripts are presented in annex 6.

5.1. Manuscript one: The cognitive-affective structure of "icyizere" (confidence) and "ishyaka" (motivation) among Rwandan students. A mixed-methods study.

Abstract

In this study, we qualitatively explored if the expectancy-value theory adequately summarised the cognitive-affective factors Rwandan lower-secondary students associated with the Kinyarwanda terms of confidence (*icyizere*) and motivation (*ishyaka*). Overall, the theory fitted the descriptions well. Compared to more recent (Western) theorising, however, we found Rwandans' value and cost beliefs to be less multi-dimensional. Utility value also seemed to be mainly about school-related aspects, something we considered to be reflective of the Rwandan socio-cultural context around education. We also established the validity of *icyizere* and *ishyaka* as translations of expectancy and value beliefs into Kinyarwanda, Rwanda's national language. We found that associating school subjects with confidence and motivation predicted school grades in these subjects.

Keywords: Expectancy-value theory; Secondary education; Africa; Rwanda; qualitative content analysis; validity

5.1.1. Introduction

Expectancy-value theory (e.g., Eccles et al., 1983) is a well-established way of conceptualising the cognitive-affective factors that drive students' engagement with learning. At the very core, motivational theory postulates that students' engagement with learning challenges is determined by the extent to which they expect engagement will lead to success (expectancy beliefs) and by the extent to which they perceive these learning challenges as enjoyable, useful, and important (value beliefs). The way students construe expectancy and value beliefs is assumed to be the product of culture (Eccles et al., 2004). This implies the notion that the way these constructs are construed also varies across cultures. It is therefore warranted to study manifestations of expectancy and value beliefs within as many different (cultural) contexts as possible (Eccles et al., 2004). Only that way can the theory's interaction with culture be traced. However, geographical distributions of expectancy-value studies (e.g., Kriegbaum et al., 2018) suggest that the theory is hardly researched within African contexts. Within the present study, we explored if the expectancy-value theory represented an adequate summary of those cognitive-affective aspects Rwandan secondary students described as part of their learning experiences. To obtain a broad range of descriptions, we used open-ended questionnaires in Kinyarwanda to ask students about their learning experiences. To obtain descriptions especially of the cognitive-affective aspects of learning, the open-ended questionnaires asked students about their experiences of *icyizere* (confidence) and *ishyaka* (motivation). We considered these terms to adequately represent the constructs of expectancy and value beliefs in Kinyarwanda, Rwanda's national language. We then applied the expectancy-value theory to students' descriptions to determine if it adequately summarised students' cognitive-affective learning experiences. This also allowed us to explore the way in which expectancy and value beliefs manifested withing the Rwandan context. The credibility of the analytical results depended upon the extent to which the Kinyarwanda terms *icyizere* (confidence) and *ishyaka* (motivation) adequately captured the meaning we ascribed to the theoretical constructs of expectancy and value beliefs. This is ultimately a question about validity (e.g., Krippendorff, 2004; Schreier, 2012). An integral part of the present study was therefore about establishing validity of our wording strategy. To do so, we carried out semi-structured interviews and correlational analysis. The latter linked school grades to the extent to which Rwandan students associated school grades with confidence and motivation. Emerging correlational patterns were then also taken as first evidence for whether or not cognitive-affective factors may explain learning outcomes in Rwanda, something the country has been struggling with (e.g., Burdett & James, 2018).

5.1.1.1. A background in expectancy-value theory

Expectancy beliefs are about how well one will perform academically within a given learning challenge. Value beliefs are about how enjoyable, useful, and important these challenges are perceived to be (e.g., Eccles et al., 1983). Within expectancy-value research, value beliefs especially have increasingly been differentiated out. According to Gaspard's et al. (2015) eight-dimensional conceptualisation, value beliefs concern the enjoyment one associates with a given subject (intrinsic value), the importance of doing will in that subject as well as the subject's importance for one's identity (importance of achievement/personal importance), the subject's usefulness for general school success (utility for school), its helpfulness for daily activities (utility for daily life), its helpfulness for getting along with friends and classmates (social utility), its helpfulness for one's general life plans (general utility for future life), and the subject's usefulness for one's professional career (job utility). More recent iterations of the expectancy-value theory have also emphasised cost beliefs as a (detrimental) determinant of engagement. Cost beliefs refers to what students feel they have to invest mentally and emotionally while being academically engaged (Eccles et al., 1983). According to Flake et al.'s (2015) work, costs is a four-dimensional construct consisting of perceptions about how challenging (effort costs) as well as psychologically taxing (emotional cost) an academic challenge is, about how many other opportunities need to be given up in order to work on that challenge (loss of valued alternatives), and about how much time and work other competing challenges require (outside-effort costs; Flake et al., 2015). Using this fine-grained conceptualisation of its different belief types, we explored to what degree the expectancy-value theory adequately summarised the different cognitive-affective experiences Rwandan students associated with learning. To collect student descriptions, we employed qualitative data-collection methods, something that seemed to be done less frequently within African classroom contexts.

5.1.1.2. The need for more qualitative research within African classroom contexts

Contemplating the different elements of the theory, Eccles and colleagues (2004, p. 169) predicted "cultural differences in success expectations and subjective task value-related beliefs". If so, then it matters that Africa is generally underrepresented within expectancy-value research. It implies that a large fraction of the worldwide student population is excluded from respective research. Within Kriegbaum's et al. (2018) meta-analysis on the predictive power of expectancy and value beliefs as well as intelligence for academic achievement, for example, only one out of the 74 quantitative studies included was from Africa (Nigeria). Within the Western world, one also increasingly finds qualitative studies on the various aspects of expectancy and value beliefs within classroom contexts (e.g., Dobie, 2019; Erickson, 2019). It is especially through qualitative methods that local manifestations of expectancy and value beliefs can be adequately explored (Eccles et al., 2004; Zusho & Clayton, 2011). In Africa, qualitative research appears to be especially concerned with "why individual children […] go to school" (Marshall, 2016, p. 423; Posti-Ahokas & Palojoki, 2014), a question generally addressed outside expectancy-value frameworks. Qualitative studies on African students' cognitive-affective experiences of learning within classroom contexts therefore appear to be warranted. Expectancy-value theory can help to guide these studies. From African viewpoints, the theory may be seen as an outside theory coming from the Western world (Eccles et al., 2004). Within the Western world, however, expectancy-value theory is a well-established model of motivation and as such can help to compare local manifestations of cognitive-affective experiences within African contexts with the way these experiences are discussed within the (Western) literature (Chao & Lambert, 2013; Zusho & Clayton, 2011). The feasibility of these comparisons, however, depends on the way these constructs are explored qualitatively within research. It is important to ensure that the way respondents construe the constructs of expectancy and value beliefs are in fact in line with the way researchers understand their meaning. Ultimately, this is a question of validity (e.g., Krippendorff, 2004; Schreier, 2012). Only to the extent that the data collected contain the information researchers wanted to obtain in the first place can they use the data to answer the study questions. Within the present study, we used the expectancy-value theory to explore Rwandan students' cognitive-affective learning experiences. During study implementation and analysis, we therefore also focused on the validity of the way we translated the study's key concepts of expectancy and value beliefs into Kinyarwanda, Rwanda's national language.

5.1.1.3. Rwanda's educational context

Rwanda is a low-income country in East Africa with Kinyarwanda as the national language. One of the visions of the Government of Rwanda has been to transform the country into a knowledge-based economy (MINECOFIN, 2000). In the wake of this, primary school fees were abolished in 2003, nine years of free basic education (9YBE) were introduced in

2008 followed by 12 years of free basic education (12YBE) in 2012. Students now pass through six years of mandatory primary education (P1 – P6) before moving on to three years of lower (S1 – S3) and three years of optional upper (S4 – S6) secondary education. Within Rwanda, the introduction of free basic education is generally considered a success. It has made Rwanda one of the top-performing low-income countries in terms of getting children into schools (UNESCO, 2014). To absorb the increase in student numbers also within secondary education, the Rwandan government introduced basic-education schools (Honeyman, 2015b; Williams, 2019). Even though not entirely free of charge (Williams et al., 2015), seeking education at basic-education schools is substantially cheaper than at boarding schools, which have traditionally constituted Rwanda's post-primary institutions (Williams, 2019). Whereas boarding schools reserve the right to choose students that performed well during national examinations, basic-education schools are free for everyone irrespective of their former performance, which explains why the latter have often been dubbed "schools for those who failed" (Williams, 2019, p. 646). This assessment may not necessarily be a fair one. For some families, basic-education schools are the only viable option for their children because they are just not able to pay tuition fees associated with boarding schools. In the present study, we recruited students from basic-education schools because these institutions accommodate the vast majority of secondary students in Rwanda (MINEDUC, 2019).

5.1.1.4. The present study

Within the present study, our goal was to explore whether the expectancy-value theory constituted an adequate framework to conceptualise the cognitive-affective factors Rwandan secondary students associate with their classroom-based learning. We were especially interested in the degree to which the different dimensions of expectancy and especially value and cost beliefs emerged within the cognitive-affective learning experiences of Rwandan students as well. We were also interested in any potential cognitive-affective aspects of Rwandan students' experiences that did not feature within the expectancy-value theory. Combining both research interests, we therefore adopted the following research question to guide the present study: *To what extent do the different cognitive-affective dimensions specified by the expectancy-value theory overlap with the cognitive-affective aspects that Rwandan secondary students associate with learning?*

We considered the credibility of our study results to hinge on our ability to make students focus on the cognitive-affective aspects of learning when describing their learning experiences. To achieve this, the wording of questions about students' learning experiences was key. Only when the wording of the questions made students reflect about their cognitive-affective learning experiences were we collecting the data we required. Thus, the present study was not only about the applicability of the expectancy-value theory to student's descriptions about their cognitive-affective experiences around learning. Rather, an integral part of the study was also about demonstrating the validity of the wording strategy chosen given our interest in student descriptions of their cognitive-affective experiences in the first place.

5.1.2. Methods

Data collection took place in May 2019 during the second of three terms that make up the Rwandan school year. To collect students' descriptions of their learning experiences, we administered open-ended, self-administered questionnaires. Especially within adult-dominated societies such as Rwanda (Adekunle, 2007), young people in rural areas are often not used to speaking up freely in front of adult strangers (Punch, 2002). They may easily perceive these interviews as "interrogations" (McWilliam et al., 2009, p. 70) and thus not be willing to open up. In these instances, the anonymity of a self-administered questionnaire can be "liberating" (Hill, 1997, p. 174) and help children to express their thoughts and feelings more openly. The effectiveness of using self-administered questionnaires has been documented in Western and low-income countries alike (e.g., Clayton & Zusho, 2016; R. B. King & McInerney, 2019; M. Lee & Bong, 2016; Posti-Ahokas & Palojoki, 2014). Within questionnaires, we also asked students to list the subjects they typically associate with (lack of) confidence and (lack of) motivation. Around five weeks after questionnaire administration, in-person follow-up interviews were carried out to explore students' understanding of the Kinyarwanda terms used within the questionnaires and to seek clarifications about some points students had raised in their questionnaire answers. However, we eventually perceived the additional value of interviews for the exploration of the themes that had emerged within the questionnaire data as rather scant.[14] Since exam periods were approaching and additional data collection could have constituted a further burden for students, we therefore decided not to continue qualitative data collection after a first round of 17 interviews. Upon completion of the 2019 school year, we also collected school grades for each study participant from participating schools.

5.1.2.1. Participants

For this study, two basic-education schools outside Rwanda's capital city (Kigali) were selected. We invited lower-secondary classes (S1 to S3) to take part in this research. Prior to data collection, the first and last author visited all target classes to explain the study

[14] We considered the fact that in-person interviews did not result in additional useful data as a confirmation of our assumption that Rwandan students may not feel comfortable to speak up freely during in-person interviews. We therefore concluded that the use of open-ended questionnaires as the main data-collection tool was indeed the right approach to the study of students' cognitive-affective experiences, at least within the confines of the present study.

purpose, and the rights of participants during study implementation. Study participation was conditional upon the completion of a parental consent and child assent form. During data collection, 170 students participated altogether. At School A, it was 74 out of all 232 lower-secondary students (32.8per cent) whereas at school B it was 96 out of 161 students (59.6 per cent). Participants were 15.27 years old on average ($SD = 1.77$). Respondents at school A ($M = 14.88; SD = 1.64$) were slightly younger than those at school B ($M = 15.57; SD = 1.82$). Around 55.3 per cent of all participants were female, a ratio rather consistent across both schools. 22 participants (13 males and nine females) took part in the in-person interviews five weeks after the main study. Upon completion of the 2019 school year, schools were recontacted to retrieve school marks for all study participants.

5.1.2.2. Instruments

Open-ended questionnaires were handed out to students only in the absence of teachers (e.g., during breaks, lessons cancelled).[15] We employed two types of questionnaires: the "confidence questionnaire" and the "motivation questionnaire". The former was about thoughts/feelings associated with confidence (*icyizere*) and represented the model's expectancy component (Wigfield et al., 2015). The latter was about the perceived reasons for experiencing motivation (*ishyaka*), which we considered to represent the value component within Eccles et al.'s model (Wigfield et al., 2015).To capture the cost component of the expectancy-value theory (Eccles et al., 1983), we devised a second version for each questionnaire type. The seconds versions were about perceived reasons for experiencing *lack of motivation* and thoughts and feelings associated with *lack of confidence*. Hence, altogether

[15] A review of the intrapersonal data suggests that in around 13.3 per cent of lessons, teachers did not show up for class. In this respect, Rwanda does not seem to be a special case. Teacher absenteeism is not unusual within the African context (Chaudhury et al., 2006)'. Also see page 94.

two sets of questionnaires each consisting of two forms (i.e., motivation/lack of motivation and confidence/lack of confidence) were administered to students. Each student completed only two forms, which altogether took around 20 minutes on average. Students completed forms either on motivation and lack of confidence or on confidence and lack of motivation. The assignment of questionnaire pairs to students was random. Each questionnaire started with some introductory sentences reassuring students that it is not uncommon to feel (a lack of) motivation or to experience (a lack of) confidence at least in some school domains. The forms then proceeded with the main question (e.g., "However, what are the reasons that we are particularly motivated about some school subjects at school?" or "Can you help us to find out what confident students may think when they are studying in class?"). In the answer field, students were then expected to provide their answers by completing sentences in relation to the questionnaire topic (e.g., "I am motivated to work on an exercise because…"). Before doing so, students first specified their age, gender, and listed subjects they associated with (lack of) confidence or (lack of) motivation. The four forms are presented in annex 2.

Table 6: The confidence coding scheme with frequencies of occurrence (manuscript one)

Category/ Sub-Category	Frequency	
	% of data segments	% of respondents
Expectancy perceptions	43.7	85.7
Effort	2.7	14.9
Affective reactions*	12.0	51.2
Negative affective reactions*	4.7	20.2
Positive affective reactions*	7.3	31.0
Consequence*	34.0	82.1
Source*	7.6	33.3

Note. *) Data-driven Coding Schemes that were added during coding. '% of data segments' refers to the percentage of all data segments that were coded using a particular coding category. '% of respondents' refers to the percentage of study participants that provided at least one answer that was coded using a particular coding category.

5.1.2.3. Qualitative content analysis

Data collection resulted in two data sets ((lack of) confidence and (lack of) motivation). They were analysed through qualitative content analysis, a method about finding descriptive categories that adequately summarise emerging themes within a text (e.g., Schreier, 2012). All data were coded by the first and second author, a male non-Rwandan and a female Rwandan. During coding, both coders principally worked with the Kinyarwanda raw data. English translations to be used in this manuscript were provided by an external translator. Each data set was analysed through a separate coding frame. The motivation coding frame was derived from expectancy-value theory and thus consisted of the expectancy and value aspects outlined in section 1.1. We also added a category on personal goals. The motivation questionnaires we used were about the "reasons" for (not) being motivated, which is often the way to operationalise (achievement) goals (Zusho & Clayton, 2011). We therefore expected statements to partly reflect students' personal goals as well. During data work, we also identified student answers that we interpreted to be about perceived learning utility and uselessness. We therefore added these two data-driven coding categories to the motivation frame along with a miscellaneous option (see table 7). As will be discussed below, however, we considered learning utility as a conceptually problematic category. The initial coding scheme for the confidence data consisted of an expectancy and task-difficulty category but was eventually extended by two categories on affective reactions associated with (lack of) confidence (see table 6). They are the equivalent to intrinsic value and emotional costs within the motivation coding frame. Unlike in the case of value and costs, we expected emotions to be manifestations rather than triggers of (lack of) confidence. To highlight this aspect, we called both categories "affective reactions" (e.g., Pinxten et al., 2014, p. 153). We also added two additional categories about the perceived sources and consequences of both (lack of) confidence and (lack of) motivation to both coding frames (for an analysis of the data within these categories see annex 7).

Table 7: The motivation coding scheme with frequencies of occurrence (manuscript one)

Category/ Sub-Category	Frequency	
	% of data segments	% of respondents
Expectancy perceptions	25.6	71.7
Cost perceptions	11.8	36.1
Effort costs	4.7	17.5
Emotional cost	6.2	23.5
Opportunity costs	0.4	3.0
Outside-effort costs	0.4	1.2
Value perceptions	31.8	57.2
Attainment value	0.0	0.0
Intrinsic value	10.3	29.5
Utility Value	21.4	49.4
Utility for daily life	0.8	4.8
General utility for future life	2.0	9.6
Utility for school	6.9	25.3
Social utility	1.4	7.2
Learning utility*	6.0	19.9
Perceived uselessness*	1.9	8.4
Utility (other)*	2.4	12.7
Goals	7.1	18.7
Consequence*	8.6	38.0
Source*	15.3	42.8

Note. *) Data-driven coding schemes that were added during coding. '% of data segments' refers to the percentage of all data segments that were coded using a particular coding category. '% of respondents' refers to the percentage of study participants that provided at least one answer that was coded using a particular coding category.

Data coding itself was done in rounds. The first two rounds served to clean the data. First, tautological student answers (e.g., "I am motivated to work on an exercise because…I have motivation in my studies") were identified and subsequently removed. Second, the overall type of each data segment was determined. At times, students provided answers to a question the corresponding form was not about (e.g., providing reasons for motivation in response to a demotivation questionnaire). These segments were not dismissed but reassigned to ensure all student statements were comparable. With the third round, the main coding work started. Given the hierarchical nature especially of the motivation coding frame, we coded the data in rounds as well (see section S5 of the supplemental material for more information). We first applied the main categories (e.g., utility value), and subsequently the sub-categories (e.g., utility for daily life). We considered the relative frequency of a coding category (see columns on the far right of tables 6 and 7) as an indication of the salience of the motivational belief associated with that category within Rwandan students' cognitive-affective experiences. Accordingly, the more percentages a coding category accumulated, the more relevant we considered the corresponding dimension to be for the manifestation of expectancy, value, and cost beliefs within the Rwandan context.

5.1.2.4. Validity

The quality of our study results ultimately depended on the validity[16] of the data collected given our research objective. A necessary but insufficient condition for validity is reliability. For each round of coding, we calculated Krippendorff's alpha (Krippendorff, 2004; Krippendorff & Craggs, 2016). In case of coding disagreements, both coders reverted back to Kinyarwanda data segments concerned and discussed them until a shared interpretation was reached (i.e., "negotiated agreement"; J. L. Campbell et al., 2013, p. 305). As for validity, we took into account its four different sources. To our knowledge, however, no (quantitative) indicators to measure expectancy and value beliefs within the Rwandan context have ever been established. It was therefore not possible to establish whether or not any of our coding categories predicted other indicators of motivational beliefs (criterion validity). Given that the coding frames were derived from the wider motivational literature and were allowed to be expanded by data-driven coding categories, we concluded that the

[16] Within qualitative research, research is often discussed in terms of its credibility and trustworthiness. Even though it also reflects a constructivist epistemology, qualitative content analysis is discussed in terms of validity and reliability nonetheless (Drisko & Maschi, 2016).

coding frames should have adequately covered all relevant aspects of Rwandan students' expectancy and value beliefs (content validity). In addition, we interviewed two groups of Rwandan students about the meaning of *icyizere* (confidence) and *ishyaka* (motivation). First, prior to data collection, we piloted the instruments using a group of five Rwandan university students (mean age of around 23). Second, around five weeks after the administration of questionnaires, we carried out semi-structured interviews with 17 of our study participants. Within interviews, we were interested in whether or not they express an understanding of both terms that reflected the way in which we conceptualised the theoretical constructs of expectancy and value beliefs (face validity). We expected students to describe the meaning of expectancy beliefs along the lines of "Can I do the task?" and value along the lines of "Do I want to do the task?" (Wigfield et al., 2015, p. 657). Lastly, we established whether or not the way Rwandan students used the terms of *icyizere* (confidence) and *ishyaka* (motivation) predicted school grades (construct validity). To do so, we asked them to specify the school subjects they associate with the (lack of) confidence) and (lack of) motivation at the beginning of any of the four questionnaires. Through biserial correlations, we then examined whether or not associating a given school subject with any of the four experiential states predicted end-of-year school grades in that subject. Accordingly, we expected confidence and motivation to positively predict and lack of confidence and motivation to negatively predict school grades. We did not expect coding categories to predict school grades. Whereas coding categories were located at the situational level, school grades were subject-specific (Swann et al., 2007). The results section starts by reflecting upon the of validity of our data.

5.1.3. Results

5.1.3.1. The meaning of the terms "icyizere" and "ishyaka"

We established face validity through the extent to which Rwandan students understood *icyizere* (confidence) and *ishyaka* (motivation) in a way we had conceptualised the constructs that stand behind these terms. To do so, we first interviewed five university students about their comprehension of both terms during pre-tests of the different open-ended questionnaires prior to the main data collection. Their explanations suggest that the two Kinyarwanda terms carry meanings that mirror their English equivalents. As one of the students explained, for example, *icyizere* would imply not having any doubts about achieving something. *Ishyaka* was described as wanting to do something because of the joy it brings or because of the instrumental value it carried. We considered the explanations of both terms by university students to be in line with the concepts behind expectancy and value beliefs (Wigfield et al., 2015), which in turn was the first evidence to support face validity of our wording strategy.

The second source to establish face validity were the follow-up interviews with 17 study participants some weeks after they had filled out the open-ended questionnaires. According to the secondary students interviewed, confidence implies "that you are able to do something" (male S2 student) or "feeling like you are going to succeed" (male S1 student), something we considered as describing the essence of expectancy beliefs (i.e., "Can I do the task?"; Wigfield et al., 2015, p. 657). Motivation was generally described as behavioural or emotional engagement. Examples for the former included statements such as motivation means “to put in effort” (male S1 student) or “doing [a task] with zeal” (female S2 student).[17] Examples for the latter included statements such as “motivation […] is to do an activity that you love” (female S1 student). In this respect, it resembled our conceptualisation of motivation (i.e., “Do I want to do the task?”; Wigfield et al., 2015, p. 657). Hence, students’ understanding of the terms “confidence” and “motivation” at least partly resembles the conceptualisation behind expectancy and value beliefs (Wigfield et al., 2015), which we took as further evidence in support of face validity of our wording strategy,

Table 8: The perceived (lack of) confidence and motivation across subjects (manuscript one)

Subject	Confidence	Lack of confidence	Motivation	Lack of motivation	Average grades
Biology	40.0 (49.3)	21.2 (41.2)	50.0 (50.3)	22.0 (41.6)	46.1 (11.4)
Chemistry	32.2 (47.0)	26.2 (44.3)	24.4 (43.2)	34.1 (47.7)	46.8 (10.6)
English	43.3 (49.8)	16.2 (37.1)	41.0 (49.5)	20.9 (40.9)	46.9 (12.8)
Entrepreneurship	14.4 (35.4)	16.2 (37.1)	9.0 (28.8)	11.0 (31.4)	47.6 (10.7)
French	4.4 (20.7)	25.0 (43.6)	5.1 (22.2)	16.5 (37.3)	50.8 (10.7)
Geography	37.8 (48.8)	21.2 (41.2)	28.2 (45.3)	30.8 (46.4)	46.2 (11.3)
History	24.4 (43.2)	26.2 (44.3)	23.1 (42.4)	18.7 (39.2)	42.6 (12.4)
ICT	11.1 (31.6)	26.2 (44.3)	10.3 (30.5)	12.1 (32.8)	66.8 (12.7)
Kinyarwanda	52.2 (50.2)	20.0 (40.3)	50.0 (50.3)	24.2 (43.1)	53.0 (12.1)
Kiswahili	15.6 (36.4)	13.8 (34.7)	10.3 (30.5)	12.1 (32.8)	51.2 (12.5)
Literature	5.6 (23.0)	0.00 (00.0)	7.7 (26.8)	7.7 (26.8)	56.8 (12.2)
Mathematics	41.1 (49.5)	37.5 (48.7)	33.3 (47.4)	34.1 (47.7)	43.5 (13.2)
Physics	34.4 (47.8)	40.0 (49.3)	24.4 (43.2)	37.4 (48.6)	52.1 (11.5)
Religion	2.2 (14.8)	2.5 (15.7)	5.1 (22.2)	7.7 (26.8)	66.5 (14.0)
Sports	2.2 (14.8)	2.5 (15.7)	9.0 (28.8)	1.1 (10.5)	82.0 (14.7)
n	90	78	78	88	170

Note. Students were only asked about either what subjects they associate with motivation and lack of confidence or what subjects they associate with confidence and lack of motivation. School grades are presented as percentages ranging from 0 (worst) to 100 per cent (best). Standard errors in brackets.

[17] A number of student answers captured through the open-ended questionnaires also referred to motivation as behavioural engagement. These associations are further discussed in section S6 of the supplemental material.

5.1.3.2. The relations between school grades and "icyizere" as well as "ishyaka"

Table 8 presents the extent to which each subject was associated with (lack of) confidence and motivation. Students linked especially the quantitative subjects (i.e., Physics, mathematics, and Chemistry) with lack of both motivation and confidence. English and Kinyarwanda were linked with both confidence and motivation. Subjects such as Entrepreneurship, Kiswahili, Religion, or Sports were generally not linked with any of the four psychological experiences in particular. Table 8 also presents average school grades within the different school subjects. Students tended to score slightly higher in the language than in the quantitative domain. Table 9 on next page presents the results of the biserial-correlation analysis combining the subject-listing and the school-grade data. On average across the school subjects, confidence and especially motivation exhibited stronger associations with school grades than lack of motivation and especially lack of confidence. For example, none of the relations involving lack of confidence were statistically significant. Correlations tended to be stronger especially within the quantitative domains. Within mathematics and Physics, both confidence and motivation positively predicted school grades. In mathematics, not only motivation but also the lack thereof predicted school grades, a somewhat contradictory finding we also observed in Geography. Within subjects outside the quantitative domain, correlations tended to be insignificant. Within the language domain, for example, it was only in the case of French and Kiswahili that we found school grades positively predicted by beliefs (i.e., motivation). Nonetheless, we considered that average correlations with school grades tended to be stronger in the case of confidence and especially motivation than in the case of lack of confidence and motivation to be in line with theoretical predictions (e.g., Eccles et al., 1983). We therefore interpreted these patterns as evidence for the construct validity of our wording strategy chosen. Given evidence for both content, face, and construct validity, we felt confident about out coding results. They are presented next.

Table 9: Bi-serial correlations (manuscript one)

Subject	Confidence	Lack of confidence	Motivation	Lack of motivation
Biology	0.09	-0.13	-0.11	0.1
Chemistry	0.05	-0.05	0.24*	-0.1
English	0.1	0	0.21	0.14
Entrepreneurship	0.24*	0.04	0.02	0.09
French	0.17	-0.08	0.3**	0.13
Geography	0.07	0.03	0.25*	0.27**
History	-0.03	-0.08	0.09	-0.09
ICT	0.13	0.06	0.03	0.04
Kinyarwanda	0.17	-0.11	0.08	0.03
Kiswahili	0.03	0.11	0.32**	-0.03
Literature	0.13	NA	0.1	-0.02
Mathematics	0.43***	-0.19	0.44***	0.22*
Physics	0.33***	-0.15	0.45***	0.04
Religion	-0.16	-0.04	0.22*	0.04
Sports	-0.05	-0.08	-0.04	-0.07
Mean	0.11	-0.05	0.17	0.05

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

5.1.3.3. Qualitative content analysis

From the 170 respondents, 321 (i.e., 86 confidence, 86 demotivation, 76 motivation, and 73 unconfidence) questionnaires were collected. Across all questionnaires, 2243 unique student answers were recorded. These answers carried 2635 meaning-bearing data segments altogether. Out of those, 220 answers were excluded especially during the first 2 coding rounds as explained above leaving us with 2022 data segments to code (i.e., 811 for confidence, 511 for demotivation, 635 for motivation, and 458 for unconfidence). The coding results for each of the main categories are discussed below. Table 6 and 7 present the frequencies of each coding category determined at the level of data segments and the level of study participants. Initial coding of the data resulted in interrater agreement rates that exhibited large variation across coding rounds and data sets. It often was also low. As section S5 of the supplemental material shows, agreement ranged from .39 to .93. As section S5 also highlights, however, in all cases of initial disagreements, a shared interpretation was eventually achieved by both coders, which we took as evidence for the intersubjectivity of the coding results presented next.

Expectancy beliefs

Within both (lack of) motivation and (lack of) confidence, expectancy beliefs constituted the category most frequently applied. Around 71.7% of all those that submitted a (lack of) motivation questionnaire gave at least one statement pertaining to expectancy beliefs. Around 25.6% of all (lack of) motivation statements were coded that way, which made it the second most common main category after value perceptions. A male S1 student wrote that “I don't have motivation when I do exercises that I do not understand”. In a similar vein, a female S3 student stated that “when I am doing my work with motivation, it makes me so excited because I feel that I have the knowledge about what I am doing”. Hence, perceived expectancies predict motivation whereas perceived lack of competence is related to demotivation. However, demotivation seems to be more related to (the lack of) expectancy beliefs than motivation. Within the former, around 33.9% of statements related to expectancy beliefs compared to 18.9% of statements within the latter. As expected, expectancy beliefs constituted the most common coding category (43.8%) applied within the confidence data. After all, confidence is just a laymen’s term for expectancy beliefs. This may be best illustrated by the following statement provided by a male S2 student: “I will succeed, and I then have very good marks which is what gives me confidence”.

Value beliefs

Overall, around 31.8% of data segments were coded as pertaining to value beliefs. Value beliefs were brought up by around 57.2 of (de-)motivation respondents. Amongst value beliefs, utility perceptions were most common accounting for around 21.4% of all data segments. It was mentioned at least once by approximately 49.4% of motivation respondents. Utility value was also the only category that was eventually enriched through three additional sub-categories (i.e., learning utility; perceived uselessness, and miscellaneous; see table 7). With about 6.9% of all coding, school utility formed the most common utility facet. Around 25.3% of respective respondents mentioned it. Students stated to gain motivation for schoolwork from the prospect of getting “good grades” (female S1 student). Others mentioned feeling demotivated about school topics “that we will not be asked in the national exam” (make S2 student). The new coding category learning utility accounted for around 6% of all motivation coding. We defined it as the perceived usefulness of a lesson or an exercise because it helps ‘to have knowledge’ (female S1 student) or to ‘learn what I didn’t know’ (male S3 student). Other utility aspects were much less common. Since both categories were hardly applied and conceptually related (Gaspard et al., 2015), we decided to merge job utility into general utility. The joint category was associated with around 9.6% of (lack of) motivation respondents but only accounted for about 2% of the segment coding. The second data-driven coding category of perceived uselessness was only found within the demotivation data set. There, it accounted for around 4.3% of all data segments and was brought up by around 16.3% of those 86 demotivation respondents. We defined perceived uselessness as the perception that exercises, lessons, or subjects are not only without any daily, general, school, or social utility but in fact to be without any use at all. For example, a male S2 student explained “I feel like that subject [i.e., mathematics or English] won’t help me with anything” (male S3 student). Talking about entrepreneurship and geography, a female S2 student stated that she felt demotivated about any of the two subjects “because I don't care about it”. Social and daily utility were both rather uncommon with about 1.4% and .8%, respectively, of the segment codings. We also found some other utility-related themes in the data such as the prospects of immediate rewards. They were rather uncommon and thus were included in the miscellaneous option. Affective factors were the only value categories that appeared in both the motivation and confidence data. Around 29.5% of those that submitted a (lack of) motivation questionnaire provided at least one statement that was coded as intrinsic value. Around 10.4% of all motivation data segments were coded that way. The most common emotional experiences associated with intrinsic value was ‘being happy about’ (i.e., *“kwishima”*; 35.5%) or ‘loving’ (i.e., *“gukunda”*; 26.5%) an activity/ subject. Within the confidence data, we coded around 7.3% of student answers as positive affective reactions. Around 31% of respondents mentioned positive reactions in response to confidence. For example, a female S2 student pointed out that confidence “makes us love studying”. Within the motivation data, we did not apply the coding category of attainment value at all. Below, we discuss reasons for why it might not have occurred in the data.

Cost beliefs

Like value, costs constituted a hierarchical category, too. Effort appeared in both the motivation (“because it's easy”; male S2 student) and demotivation data (e.g., “when they have given me a difficult exercise or questions I don't understand […] I cannot get motivation”; female S3 student). Effort was associated with (lack of) motivation by around 17.5% of respective respondents and accounted for 4.7% of all de(motivation) segments. As expected, it also appeared within the (lack of) confidence data. As a female S3 student explained “the subject in which we are given the exercise is so difficult I have no confidence of ever understanding that subject”. Effort perceptions were brought up by around 14.9% of (lack of) confidence respondents but only accounted for about 2.7% of segments there. Within motivation, emotional costs only appeared within the demotivation subset where it was mentioned by around 45.8% of respective respondents and accounted for about 6.2% of codings. To a large extent, emotional costs were expressed as the absence of positive emotions (e.g., “I don't like certain subjects”; male S3 student) rather than actual negative emotions. Amongst the latter, hate appears to be the most common one. It was used in around 11.9% of respective student answers. As for unconfidence, negative affective reactions were mentioned by around 39.5% of respective respondents. It accounted for around 13.2% of their answer statements. Unlike emotional costs, however, negative affective reactions were expressed not only through the absence of positive emotions. Students referred to them through specific emotions such as sadness, hate, discouragement, fear. For example, a female S2 student wrote “when I am not confident, I feel so sad because in life confidence is everything”. As for the two final aspects of costs, opportunity as well as outside-effort costs, they do not seem to matter much in the Rwandan context. Combined, they only account for less than 0.9% of student answers coded within the motivation data.

Personal goals

The motivational category of Goals was distinguished from utility based on verb phrases used. The former are identified through phrases such as “I want to” or “I have to” (Hulleman, Schrager, et al., 2010) whereas the latter is communicated through statements such as “it makes me”, “it helps me”, or “so that I can”. Goals were associated with motivation by around 18.7% of respective respondents and accounted for around 12.6% of motivation codings. As one should have expected, goals again materialized as a hierarchical category. Students mentioned a variety of different goals they pursued through school including social-status goals (e.g., “I want to be a powerful person”; female S2 student), social approval goals (e.g., “I want to make my parents happy”; male S3 student) but also mastery goals (e.g., “I want to be intelligent”; male S2 student). This is in line with the motivational literature on achievements goals (Dowson & McInerney, 2003). The coding results on consequence and source are discussed in detail in annex 7.

5.1.4. Discussion

Within the results section, we discussed the extent to which each of the different cognitive-affective dimensions of expectancy, value, and cost beliefs emerged within respective descriptions of learning experiences Rwandan students provided. Within the results section, we also discussed new aspects of Rwandan students’ learning experiences that emerged especially within value components of the data. The per centages each coding category accumulated, so we argue, provided then an indication of the relative importance of the corresponding dimension within the manifestation of expectancy, value, and cost beliefs within the Rwandan context. The section below summarizes these results and provides a comparison of the way in which expectancy, value, and cost beliefs manifest within the Rwandan context with how these beliefs are conceptualised within the respective literature.

5.1.4.1. Expectancy beliefs amongst Rwandan students

Even though we considered motivation to loosely represent the value components of Eccles' et al. theory, the questionnaires on (lack of) motivation did not only yield student answers on value aspects. In fact, around 71.7% of respondents that completed any of the two questionnaires provided at least one statement that was coded as pertaining to expectancy beliefs. In fact, the category on expectancy beliefs appeared significantly more often in the data than those on value (57.2%) and cost beliefs (36.1%). It suggests that expectancy beliefs may constitute important drivers of motivation. This is in line with the motivational literature on self-beliefs, which has long been highlighting the centrality of expectancy beliefs, often operationalised through academic self-concept, for motivation and learning (e.g., Marsh, Seaton, et al., 2019). Meta-analyses have also pointed out the relative importance of expectancy beliefs (self-concept) as a predictor of achievement especially when compared to other motivational constructs such as value beliefs (e.g., Kriegbaum et al., 2018; Lavrijsen et al., 2021). The latter aspect was not necessarily supported by our own correlational analysis. Quantitative associations suggested that motivation, as a proxy for value beliefs, was a stronger predictor of achievement than confidence. However, Rwandan students strongly associated motivation with expectancy beliefs as well. Hence, motivation was an umbrella term that did not only refer to value beliefs but also to expectancy (and cost beliefs). This in turn may also explain why motivation might have been a stronger predictor of school grades than confidence, which could be taken as a proxy merely for expectancy beliefs.

Whereas the motivational and correlational data provided some insights into the relative importance of expectancy beliefs, the confidence data provided some qualitative insights into their local manifestation. Initially, we conceptualised confidence as a cognitive category that may materialize in form of expectancy beliefs or effort beliefs (e.g., Eccles et al., 1983). As the analysis revealed, however, Rwandan students associated affective experiences with confidence and unconfidence as well. For example, a female S2 student expressed that “I think I can do it, and this makes me happy”. This close connection between cognitive self-evaluations and corresponding affective reactions is in line especially with self-concept research. Whereas expectancy-value theory considers expectancy beliefs and affective experiences as distinct constructs (e.g., Eccles et al., 1983), self-concept researchers have partly understood competence beliefs as an “cognitive and affective appraisal of self” (Bong & Skaalvik, 2003). Whether or not cognitive and affective experiences are indeed strongly related within competence beliefs held by Rwandan students is best studied using factor analytical approaches (e.g., Pinxten et al., 2014). Within the context of the present study, we concluded that the way in which expectancy beliefs manifest within the Rwandan context did not differ significantly from what is implied by the respective motivational literature both in terms of their relative importance as well as phenomenological structure.

5.1.4.2. Value beliefs amongst Rwandan students

Within the Rwanda context, value beliefs appeared to be multidimensional as well (e.g., Gaspard et al., 2015). Interestingly, attainment value did not seem to be relevant within the Rwandan context, which is remarkable given that it is seen one of the major elements of value beliefs (Eccles et al., 1983). However, there may be some conceptual overlaps with the new category of perceived uselessness. Whereas attainment value is about the “personal importance” (Eccles, 2005, p. 109) attached to a task, perceived uselessness implies the attitude of not caring about the task or subject at all. As such, perceived uselessness could be seen as the perceived lack of attainment value. As a concept, it may also be related to what is discussed as “useless schooling” (Ailei, 2019) within ethnographic educational research, which describes the notion that “studying is not useful, and it is better not to study” (p. 186). Utility value was the most common value aspect and as a motivational factor only second to expectancy beliefs. Intrinsic value was relevant, nonetheless. Students referred to it when describing motivation. Within utility value, utility perceptions related to outside-school aspects (e.g., daily utility, general utility, and social utility) hardly seemed to matter. None of them accumulated more than 2% of all segments. Those utility facets that were related to schooling (e.g., school and learning) appeared to be more relevant. Together, they accounted for around 60% of all segments within the utility category. The emphasis of school-related utility within student answers matched the way in which the cultural context around education in Rwanda has been described. For example, Honeyman (2015b) observed that students “did not really expect or even want their classes to help them develop creativity, problem-solving, or other practical skills. […] What counted for them, crucially, was school’s role in preparing them for examinations, the hurdle they had to pass in their objective of getting a university degree and finding a high-paying job” (p. 189). Despite its relative frequency within the data, we did not find learning utility a conceptually useful category. It was often not clear why students wanted to “have knowledge” (female S1 student). Students may want to gain knowledge for different reasons such as interest (i.e., intrinsic value) or because they want to pass an upcoming exam (i.e., school utility). The actual motivational effects of learning utility may therefore depend on those reasons behind wanting to gain knowledge (a similar argument has been made for the case of the motivational construct of outcome goals; see Grant & Dweck, 2003). To sum up, within the Rwandan context we found value beliefs to emerge especially in form of utility as well intrinsic value. Within utility value, especially those aspects seemed to be relevant that were school related.

5.1.4.3. Cost beliefs amongst Rwandan students

Cost perceptions were explored across both the datasets on lack of confidence as well as lack of motivation. Like value, the construct of cost beliefs appeared to be less multi-dimensional than the one discussed in the Western-borne literature (Flake et al., 2015; Gaspard et al., 2015). Within our data, it manifested itself mainly through effort beliefs as well as emotional costs. Within the motivational literature, the role of emotions has been increasingly discussed (e.g., Pekrun & Linnenbrink-Garcia, 2014). Emotion researchers have identified at least 16 positive and negative emotions they consider relevant for the educational domain (e.g., Pekrun & Perry, 2014). Within our own data, we found the range of emotions mentioned by Rwandan students to be more limited. For example, ‘to be happy’ (i.e., kwishima), ‘to like/ love’ (i.e., gukunda), and ‘to be pleased’ (i.e., kunezeza) seem to cover most positive emotional experience amongst Rwandan students. This, however, may be due to the idiosyncrasies of Kinyarwanda, Rwanda’s main language. Especially the verb ‘kwishima’, for example, is a polysemous word. It is used to express not only happiness but also other positive emotions such as pride. it is therefore to assume that Rwandan students experience the full range of achievement emotions as well even though they may use less words to describe it. Further research using amongst others prototype methods (e.g., Shaver et al., 1987) needs to further explore this.

5.1.4.4. The overall importance of cognitive-affective factors for Rwandan students’ learning

As discussed earlier, student answers during pre-tests as well as during follow-up interviews suggested provides evidence in support of the face validity of our wording strategy. The correlational analysis provided evidence of the construct validity of the terms of *icyizere* (confidence) and *ishyaka* (motivation) as a way of expressing cognitive-affective aspects of learning. Subjects that were associated with confidence and especially with motivation were also associated with higher school grades. This was in line with what expectancy-value theory predicted (e.g., Kriegbaum et al., 2018). However, these correlational results can be taken as evidence not only for overall construct validity of our wording strategy. As concepts, *icyizere* (confidence) and *ishyaka* (motivation) may constitute rather crude conceptualisations of students' expectancy and value beliefs. Nonetheless, they predicted school grades. This in turn may point to the general relevance of cognitive-affective factors for explaining learning outcomes not only in the Western world (Kriegbaum et al., 2018; Lavrijsen et al., 2021) but also amongst Rwandan students. Further research on the nexus between motivation and learning outcomes within the Rwandan context appears to be warranted. A more sophisticated operationalization of expectancy and value beliefs, potentially through the qualitative insights obtained through the present study, may help to better trace out the actual empirical relations between students' cognitive-affective experiences and their learning outcomes. The insights one could expect from this research agenda may not only be of academic but also of practical interest. Insights into how cognitive-affective factors such as expectancy and value beliefs relate to learning outcomes could be used to design motivational interventions. Within the Western world, these interventions have been found to be impactful while being light touch (e.g., Yeager & Walton, 2011). Especially the latter aspect may also be of interest to education policymakers in Rwanda, a country that continued to struggle with learning attainment especially amongst basic-education students (Burdett & James, 2018).

5.1.4.5. Limitations and conclusions

Some study limitations must be highlighted. First, the organization of the study around Eccles' et al. expectancy-value theory helped comparing Rwandan students' cognitive-affective experiences with those discussed in the (Western) literature. However, designing the data collection tools around the constructs of expectancy and value beliefs may have also nudged students to solely thinking along those lines. This in turn may have hampered the identification of culture-specific cognitive-affective aspects that otherwise characterize Rwandan students' learning experiences. Qualitative studies that explore phenomena of interests predominantly through local frames of references (Zusho & Clayton, 2011) may further help to determine if we indeed catch all relevant cognitive-affective drivers of Rwandan students' learning experiences. A second limitation was the sample composition. One would expect that motivational beliefs are also the outcomes of person-level characteristics such as age. One would not expect young students to ponder much about the job utility of their school subjects. However, the older students become the more eminent the question about the usefulness of a given subject vis-à-vis their career prospects may become as well. It may also be for this reason that we did not fun much relevance of general/ job utility within our data. Our sample students were partly as young as 13 years. A student sample that is older on average may thus lead to partly different results. Future replications should also control for school location (e.g., urban versus rural) and school type (i.e., boarding schools versus basic-education schools). These characteristics may also be associated with different motivational beliefs amongst students (Williams, 2019). A third limitation concerned the way data on what subjects are generally associated with (lack of) confidence and motivation. We asked students to list subjects they either associated with lack of confidence and motivation or with lack of motivation and confidence. A more effective way would have been to have students rate each of their subjects in terms of the level of confidence and motivation they typically experience. This would have doubled the amount of data available to perform the correlational analysis. Despite these limitations, we considered our study an important contribution to the study of "cultural differences in success expectations and subjective task value-related beliefs" (Eccles et al., 2004, p. 169) especially within the African context. Our analysis revealed that even though the expectancy-value theory by and large applied to the Rwandan context as well, there indeed appear to be some cultural differences between the Rwandan context and the wider motivational literature. These differences manifest especially within the value component of Eccles' et al. theory. Our data suggested that especially those value perceptions were less common amongst students that focussed on aspects unrelated to schooling (e.g., daily, general, social utility). We propose that these value perceptions indeed reflect socio-cultural attitudes around education in Rwanda.

5.2. Developing the subject- and lesson-specific motivational scales: Turning thematic and verbal data into self-report items

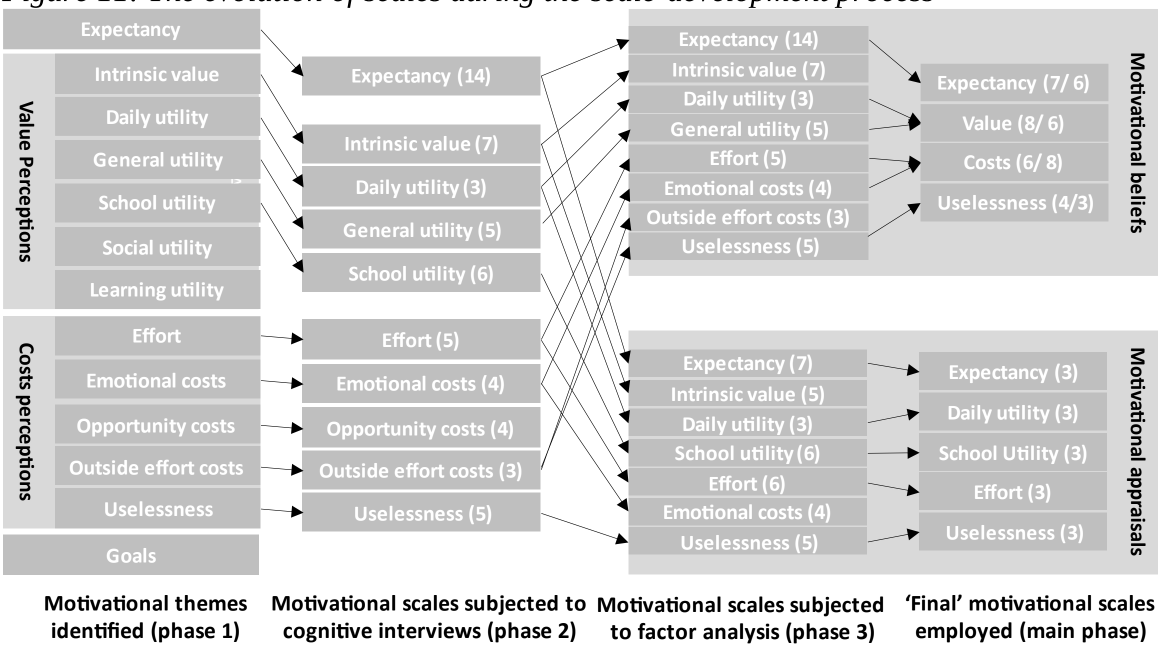
5.2.1. Introduction

Upon completion of the exploratory-qualitative phase, whose results are presented in manuscript one (see section 5.1), both the thematic as well as the raw data collected were used to derive self-report scales to measure the different dimensions of achievement motivation amongst Rwandan secondary students quantitatively. Scales were derived for both the school subject-specific and lesson-specific level. The process for each scale type was divided into a scale-development as well as scale-testing stage. For both scale types, scale development is discussed in sub-section 5.2.2. Testing of both the school subject-specific and lesson-specific scales is discussed in sub-section 5.2.3. The school subject-specific scales were tested in particular through a two separate exploratory factor analyses. The lesson-specific scales were tested through a single separate exploratory factor analysis. The extent to which the school subject-specific and lesson-specific scales differed has already been discussed in section 4.2.3 above.

5.2.2. Scale development

The starting point for developing the motivational scales was the study presented in manuscript one. Through the study, 13 motivational themes were identified. 12 of them were attributed to the different dimensions of the Eccles' et al. (e.g., 2020) expectancy-value framework. Subsequent scale-development efforts that proceeded the study reported by the Bulla et al. (2022) focussed on those 12 themes (also see figure S1 on the far left). Upon review of those 12 themes, however, it was also decided to exclude learning value and social utility from the subsequent development of scales on school subject-specific motivational beliefs. Besides perceived usefulness, learning value was the only motivational theme that emerged during the exploratory-qualitative phase that are not discussed within the mainstream literature on student motivation (e.g., Flake et al., 2015; Gaspard et al., 2015; Wigfield & Eccles, 2000). As discussed in manuscript one, there was a conceptual problem associated with learning value. It was not entirely clear why it was perceived to be useful to know what these students studied. Due to its lack of conceptual clarity, it was therefore decided against turning learning value into a motivational scale. Social utility was dropped as well. Social utility turned out to be rather multi-faceted (referring to peers, parents, and teachers). However, it was felt there was not sufficient raw data on social-utility beliefs collected to generate enough self-report items to operationalize the different dimensions of social utility. Thus, items were developed for 12 scales. At this point, it was not distinguished between the lesson-specific and school subject-specific scales.

Figure 11: The evolution of scales during the scale-development process



Note. Boxes indicate themes within motivational experiences (study 1). Numbers in brackets in this respect indicate the number of self-report items included into a given scale.

In order to develop self-report items to operationalize the motivational themes selected, raw data on the language Rwandan students used to express motivational experiences collected by Bulla et al. (2022) was re-examined. This time, of interest were especially the verb phrases students used to describe confidence and motivation. A verb phrase refers to the group of verb(s), adverbs, and objects that jointly indicate those actions or states a sentence is indented to describe (Fabb, 2005; Finegan, 2004), a pattern that also holds for Kinyarwanda (Kimenyi, 1978). Verb phrases are of importance within self-report scale development. According to Albert Bandura (2006, p. 308), for example, self-report self-efficacy items require the inclusion of the verb phrase 'I can do' in order to effectively eliciting self-efficacy appraisals. All answer statements within the raw data that were assigned to any of the nine motivational themes were therefore broken down into verbs, negation (where applicable), adverbs, and objects. A simple frequency analysis then helped determining those verb phrases that were commonly used to express a given motivational perception. The most common verbs associated with a theme were then used to devise those self-report items needed to operationalize the motivational belief associated with that theme. In this respect, auxiliary verbs students used within their descriptions were less of interest. In many instances their usage was mainly the product of colloquialism rather than grammatical necessity.

Table 10: An example of the analysis of student raw data (scale development phase)

Kinyarwanda original	English translation	Auxiliary	Verb	Negation	Object
Hari n'igihe nanga ku wu kora nawupinze kuko **mba numva ntabwo ndibushobore**.	There are times, I refuse to do it, I ignore it because **I am feeling I am not able to do it**.	Kumva	gushobora	ntabwo	bu

An example taken from the raw data analysed in the first manuscript may best illustrate the approach adopted to devise self-report items. Table 10 above presents an answer statement provided by a 16-year-old female secondary-two student. Her statement was coded as being about unconfidence (i.e., lack of expectancy beliefs). The statement consists of two parts. The first part (i.e., "there are times, I refuse to do it, I ignore it") was interpreted to be about the consequences of unconfidence. According to the student, whenever she feels unconfident, she disengages with schoolwork. The second part (i.e., "I am feeling I am not able to do it") was interpreted to exemplify a typical thought that she associates with unconfidence.[18] It was therefore coded as an expectancy cognition. This answer part was then broken down into the different elements. The verb 'gushobora' (i.e., 'to be able to') was then included into the frequency analysis performed on all answer statements pertaining to expectancy beliefs. The verbs that most commonly appeared within the raw data were then used within the self-report items on motivational beliefs. To keep items as simple as possible, each question was formed mainly through combining selected verbs as well as the subject to which the item relates. For example, two most common verbs used to express intrinsic value was 'kwishima' (i.e., to be happy about) and 'gukunda' (i.e., to love). The corresponding two self-report items therefore read "Ese [SUBJECT] riragushimisha?" (i.e., "Does [SUBJECT] make you happy?") and "Ese ukunda [SUBJECT]?" (i.e., "Do you love [SUBJECT]?"). As figure 11 highlights, following this approach resulted in 45 items across the nine scales. Out of those, seven items were worded negatively. Within the scale on perceived uselessness, all items were negatively worded. As explained in manuscript one, this construct refers to "perceptions of exercises, lessons, or subjects not only to be without any daily, general, school, or social utility but in fact to be without any use at all". It implies the negation of any perceived value within a given subject. To fully convey the absoluteness perceived uselessness implies, related questions were based on wordings involing "not". For example, one item reads:

[18] The auxiliary verb kumva (i.e., 'to feel') may suggest that the student wanted to express an affective association with unconfidence. Like in English (e.g., "I feel like I cannot do it"), however, individuals at times use this auxiliary to express an experience that is basically cognitive in its nature.

"Ese [SUBJECT] nta kamaro [ri]fite?" (i.e., Does [SUBJECT] have no value?). Within the provisional scales on expectancy beliefs as well as intrinic value there was each one negatvely worded item as well. At this stage, all items were coupled with a four-point response format (i.e., 1- yes, very much; 2- mainly yes; 3- mainly no; 4- no, not at all). These provisional scales were then subjected to the scale-testing stage.

5.2.3. Scale testing

5.2.3.1. Study 1: Pretesting provisional scales using cognitive interviews

The generation of self-report items to operationalize the preselected themes resulted in 56 provisional items across the nine scales (see figure 11 on page 143), all coupled with a four-point response format. To further evaluate the provisional scales, cognitive interviews were performed. Through cognitive interviews, researchers try to better understand respondents' comprehension of and thought processes around self-report questions (Beatty & Willis, 2007; Karabenick et al., 2007; Presser et al., 2004; Woolley et al., 2004). 10 lower-secondary students (seven females) from a public school outside Kigali participated in cognitive interviews carried out over two days towards the end of the 2019 school year. For items of all scales but one a lesson-specific and a school subject-specific version was prepared. Only school utility was available as a lesson-specific version. As explains in manuscript one, school utility is the perceived usefulness of something for one's own academic success. Both subjects and individual lessons within subjects can be perceived to be useful for one's academic success. Lesson-specific utility may emerge if, for example, a classroom exercise helps students to determine whether nor they understood a topic. By contrast, school utility at the subject level mainly referred to whether or not a school subject is high-stake subject. The underlying notion is that students tend to feel more demotivated in school subjects that are not, for example, part of national examinations. This coincides with what Honeyman (2015b) reports. She observed students to become increasingly disengaged with entrepreneurship, a subject introduced at Rwandan secondary schools some years ago, after it was announced it will not be part of national examinations. However, whether or not a subject is high stakes is common knowledge and not based on subjective perceptions. It was therefore decided against including a school subject-specific version of school utility. It exists as a lesson-specific scale only.

All scales were administered as school subject-specific questionnaires. Students either completed a mathematics or a Kinyarwanda version of each of the motivational scales. A male research assistant who was very familiar with all 10 scales was used for facilitation and translation services during the interviews. Cognitive interviews were carried out following the approach reported by Karabenick et al. (2007). Students were first asked to read out a self-report item and then to explain what it is about. Afterwards, they were requested to choose one of the four answer options best representing their self-perceptions in this respect as well as to explain why they chose it. Where deemed necessary, students were probed for additional explanations and feedback. Probing, however, was only about comprehension and interpretations of item aspects rather than the content of actual answers given. The latter may have given the impression of interviews being "interrogations" (McWilliam et al., 2009, p. 70). Upon completion of cognitive interviews, all items were reviewed.

One of the themes that emerged in the study by the Bulla et al. (2022) was about opportunity costs associated with learning. It was operationalised through self-report questions such as "Do you need to sacrifice time with your friends to study [subject]?". Cognitive interviews revealed that students interpreted these items especially to be about some form of self-regulation. For example, according to a female S2 student the above-mentioned question is about what type of friends one should have. In this respect, she explains that “you should not hang out with friends that distract you from your schoolwork”. This is not in line with the essence of opportunity costs of learning which is about what a student needs to give in order to be able to engage with learning (Flake et al., 2015). The scale on opportunity costs was therefore dropped from further scale-development efforts, which further reduced the number of scales employed to eight (also see graph S1 above). Other scales were not removed but slightly revised. For example, within the raw data used by Bulla et al. (2022) the verb ‘gutsinda’ (i.e., “to succeed/ win”) was a common way to express expectancy beliefs and thus was also used to form a respective self-report item. The cognitive interviews revealed that at least some students tend to interpret it as ‘gukunda’ (i.e., “to like/ love”), which was a surprising result. It was not clear of students just misread it or if there was any other association behind it. It was therefore decided to remove this item from the expectancy-belief scale. Hence, upon completion of the cognitive interviews 38 items across eight scales remained (also see graph S1). During cognitive interviews, the four-option response format (i.e., 1- yes, very much; 2- mainly yes; 3- mainly no; 4- no, not at all) was piloted as well. The feedback obtained by Rwandan students (and teachers) suggested a general confusion about the relation between the different answer options. For example, respondents mentioned that it was not clear what the difference is between the first and the second answer option. It was therefore decided to go a three-point answer format consisting of 1) often; 2) sometimes; and 3) never.

5.2.3.2. Study 2: Exploratory factor analysis of school subject-specific motivational scales (EFA I)

Upon completion and analyses of cognitive interviews, provisional domain-specific motivational scales were revised and then subjected to exploratory factor analysis (EFA). EFA allows assessing the internal structure of provisional scales.

Method

Participants and Procedure

Data was collected in the second half of the last term of the 2019 school year. For this preparatory study, a public school was randomly selected out of the school register of one of the Rwandan districts outside Kigali. The school is located in the district's remote parts. Its lower-secondary section comprised 698 students in total. To recruit students, all target classes were visited prior to data collection. On those visits, the purpose of the study was introduced, the data-collection procedure outlined, and rights of participants explained. Participation was voluntary and was not tied to any form of benefit. For students to participate, both child assent and parent consent forms were required. 431 students (61.75%) submitted valid consent forms and thus were included into the study.

On the day of data collection, questionnaire administration was handled by three survey teams each consisting of two enumerators recruited from the district of the school. Questionnaires were only handed out to students in their classrooms outside study time (e.g., lessons cancelled). Before handing out questionnaires, enumerators read out questionnaire instructions to students in Kinyarwanda. The purpose was to explain the different sections of the questionnaire, the logic behind the answer options, and to highlight again students' rights during the research process. During data collection not all participating students were found in class. Only 371 out of 431 students submitted valid questionnaires (86.1%). The final study sample was slightly skewed towards younger and female students. 48.8% of students attended secondary one (mean age = 14.24 years), 28.8% attended secondary two (mean age = 15.58 years), and 22.4% attended from secondary 3 (mean age = 16.1 years). Around 62% of participants were female. Data collection was paper based. To avoid any disruptions during data collection in class, questionnaires were given out to all students who wanted one. Questionnaires from students that had not submitted valid consent forms were removed afterwards. Data from all questionnaires were entered twice into the computer, each round from a different data-entry clerk. Discrepancies between both rounds were reconciled and then exported for data analysis (Paulsen et al., 2012). Cases where students provided no answer or more than just one answer were coded as missing values (-9 and -99 respectively) during data entry and subsequently removed from the data analysis.

The Provisional Motivational-Beliefs Scales

Scales on motivational beliefs was administered in two versions (i.e., two different questionnaires). Both consisted of a Kinyarwanda and a mathematics section. One version started with the Kinyarwanda and ended with mathematics whereas the other one started with mathematics and ended with Kinyarwanda. Around 51% of participants completed the former and approximately 49% completed the latter. The content of both sections was identical. Each consisted of eight scales with 38 items in total (also see section S2 above). The expectancy scale consists of eight items (e.g., “do you understand [SUBJECT]?”), the daily-utility scale consists of three items (e.g., “does [SUBJECT] help you in your daily life?”), the general-utility section consists of five items (e.g., “will [SUBJECT] help you in your future life?”), the perceived uselessness scale consists of five items (e.g., “do you think [SUBJECT] is useless for you?”), the intrinsic-value scale consists of six items (e.g., “does [SUBJECT] make you happy?”), the effort scale consists of four items (e.g., “is the subject of [subject] difficult?”), the outside-effort scale consists of three items (e.g., “are you too busy to study [SUBJECT]?”), and the emotional-cost scale consists of four items (e.g., “do you hate this [SUBJECT]?”). Based on feedback gathered during cognitive interviews (see section S2 above), the item response format was a Likert scale with three options: 1) Most of the times; 2) Sometimes; 3) Never. Questionnaires were only given out in Kinyarwanda. Questionnaire administration took around 36.7 minutes per class. Before analysis, all items were reverse coded and thus higher numbers imply higher frequency.

Exploratory Factor Analysis

Before the exploratory factor analysis was performed, summary statistics of the different scale items were obtained. For each scale item, means, standard errors, and inter-item correlations were determined. Given item’s 3-point response format, the current data was neither approximately continuous (Dolan, 1994) nor normal (Finney & DiStefano, 2013). Instead of Pearson correlations, correlations and thus factor analysis were therefore based on the polychoric correlations (Olsson, 1979), whoch is also robust to deviations from normality (Finney & DiStefano, 2013; Flora & Curran, 2004; Holgado–Tello et al., 2010). Polychoric correlations were estimated through the *polychoric* function in the *psych* package in R (Revelle, 2020) using R software package version 3.6.2 (The R Core Team, 2019).

A key step in performing exploratory factor analysis is to determine the number of factors to be extracted (Fabrigar & Wegener, 2012; Gaskin & Happell, 2014; Osborne, 2014; Schmitt, 2011). To do so, the eigenvalues of the observed polychoric correlation matrix were analysed using Horn's parallel analysis, probably the most recommended method for dimensionality assessment (Garrido et al., 2016; e.g., Henson & Roberts, 2006). Within R, parallel analysis was implemented using *fa.parallel* of the psych package (Revelle, 2020) using the 95th per centile and principal axis factoring (e.g., Finch, 2020; Garrido et al., 2013). There has been some debate about the potential of model fit indices to identify the number of factors within exploratory factor analysis (D. A. Clark & Bowles, 2018; Finch, 2020; Garrido et al., 2016; Montoya & Edwards, 2021). Within an EFA framework, indices of model fit indicate how well a factor solution imposed on a matrix matches the observed correlation matrix and that way provide additional evidence on what number of factors to retain (D. A. Clark & Bowles, 2018; Finch, 2020; Garrido et al., 2016). Common fit indices include in particular the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). Both the CFI and TLI are goodness-of-fit indices. Values of .95 or higher are seen as indicator of an excellent fit (West et al., 2012). By contrast, a RMSEA value of .06, or lower, is seen as evidence of indicative of good model fit (Hu & Bentler, 1999). However, especially overfactored models tend to exceed the aforementioned thresholds as well (D. A. Clark & Bowles, 2018). Instead of absolute cut-off values alone incremental improvements within fit indices may provide more robust evidence on model adequacy ("fit statistic difference approach"). In fact, in more unfavourable cases with small sample sizes and items to be categorical, a RMSEA difference value of 0.015 seem to perform better at determining the right number of factors than other, more traditional dimensionality-assessment methods including parallel analysis. The rule implies to abandon a more parsimonious model in favour of another model that consists of one more factor if the latter results in fit improvements within the RMSEA statistic of at least 0.015. A respective cut-off value of 0.01 for CFI and TFI has been found to perform worse (Finch, 2020).

All factor models were estimated in Mplus, version 8.7 (L. K. Muthén & Muthén, 2017). To account for the nonnormal, categorical data, all factor models were extracted from the polychoric correlation matrix using diagonally weighted least squares (WLSMV) estimation (Finney & DiStefano, 2013). Since latent factors were expected to be partly correlated (Watkins, 2018), factor extraction was implemented with oblimin (i.e., oblique) factor rotation (Fabrigar & Wegener, 2012). The various factor solutions obtained were then compared in terms of both model fit (e.g., RMSEA) as well as conceptual plausibility (Fabrigar & Wegener, 2012; Watkins, 2018). The conceptual plausibility of factor solutions was assessed through evaluating the pattern matrix (J. M. Graham et al., 2003; Watkins, 2018). Factor solutions that are plausible conceptually exhibit a “simple structure” (Brown, 2015, p. 27) within the pattern matrix. In those cases, each item in the data is correlated (i.e., load) with one and only factor (Fabrigar & Wegener, 2012; Gorsuch, 2015; Watkins, 2018). This in turn allows to easily interpret the content of a factor (Fabrigar & Wegener, 2012; Kline, 2013). Researchers generally go for the factor solution that exhibits the best balance between relative model fit and conceptual interpretability (Watkins, 2018). Items that show either only poor loadings (< 0.3) across factors or that exhibit substantial loadings (>0.3) on more than just one factor are usually revised or even removed (Fabrigar & Wegener, 2012; McCoach et al., 2014). The exploratory factor analysis was carried out separately for both domains. A sample size of 371 data points would have been too small for a joint analysis. Missing data accounted for around 1.8%. To deal with missing data, the FML approach was implemented in Mplus, in which all available information is taken into account (Asparouhov & Muthén, 2010).

Results

Descriptive Statistics

Across the scales, means for each item was high with 2.28 ($SD = 0.53$) in Kinyarwanda and 2.2 ($SD = 0.34$) in mathematics. Averages were particularly high amongst items related to value ($M's = 2.8, 2.6$) compared to expectancy ($M's = 2.53, 2.12$) and cost beliefs ($M's = 1.81, 1.94$). Value items were also more (negatively) skewed ($M's = -2.14, -1.28$) and exhibited stronger kurtosis ($M's = 7.02, 4$) compared to other items in Kinyarwanda (average skewness $= 0.03$, average kurtosis $= 2.37$) and in mathematics (average skewness $= 0.11$, average kurtosis $= 2.31$). Factor analysis based on DWLS estimation, however, is robust to non-normal data (Finney & DiStefano, 2013).

Table 11: Summary of the intra-scale correlations (scale testing phase; EFA study I)

(Provisional) Scale	Kinyarwanda			Mathematics		
	Mean[1]	Min	Max	Mean[1]	Min	Max
Expectancy[2]	0.46 (0.53)	-0.32	0.78	0.50 (0.60)	-0.30	0.85
Daily utility	0.47	0.32	0.57	0.64	0.56	0.73
Effort belief	0.49	-0.33	0.74	0.51	-0.44	0.73
General utility	0.61	0.29	0.78	0.74	0.61	0.83
Emotional costs	0.74	0.61	0.83	0.58	0.42	0.66
Outside effort	0.25	-0.18	0.50	0.36	-0.29	0.61
Perceived uselessness	0.69	0.59	0.86	0.67	0.53	0.82
Intrinsic value[2]	0.46 (0.64)	-0.18	0.75	0.46 (0.67)	-0.07	0.77

Note. [1]) mean was calculated based on absolute values of all intra-item correlations. [2]) mean in brackets refer to average intra-scale correlations between items without the negatively worded question.

Table 11 above presents a summary of the correlation matrices for self-report items. Within mathematics, correlations are slightly higher ($M = 0.55; SD = 0.24$) than in Kinyarwanda ($M = 0.52; SD = 0.22$). There are differences between scales as well. Correlations are particularly high in emotional costs (Kinyarwanda), general utility (mathematics). Across both domains, correlations are significantly lower within items on outside effort costs suggesting there are some problems with this scale. Across both domains, negatively worded items also do not seem to perform well. Initially, provisional scales on both expectancy beliefs and intrinsic value contained one negatively worded item. In both cases, these items showed only weak correlations with other scale items. As table 11 highlights, removing negatively worded items substantially increased average interitem correlations. Factor analysis was therefore performed without negatively worded items.

Table 12: Results of dimensional assessment (scale testing phase; EFA study I)

Method	Number of Factors to be retained	
	Kinyarwanda	Mathematics
Scree Plot	3	4
Horn's Parallel Analysis with 95% Per centile	7	7
UCP Adjusted Parallel Analysis	3	4
Minimum Average Partial Procedure	5	4

Exploratory Factor Analysis

Parallel analyses of the items suggested a seven-factor solution for both Kinyarwanda and mathematics domains (see table 12). For both Kinyarwanda and mathematics, factor models ranging from three to seven factors were estimated. As table 13 on page 166 shows, especially the four-factor through seven-factor solutions exhibited good fit. Comparing the different models using a RMSEA difference value of 0.15 and a CFI/ TLI difference value provides further evidence on the number of latent factors to be retained (Finch, 2020). The former suggested a four-factor solution to be most adequate within the mathematics domain. Within Kinyarwanda, a four-factor solution just barely misses a RMSEA cut-off value of 0.015. Employing a cut-off value of 0.01 for CFI/ TLI suggested a four-factor model not only within Kinyarwanda but mathematics.

Within all models, there was a number of items that exhibited substantial cross loadings. Models involving five or more factors tended to contain at least one factor that was ill defined, which suggests they may have been over-extracted. These factors were only associated with items that exhibited substantial double loading. The seven-factor model within Kinyarwanda even contained two ill-defined factors. Since it is difficult to interpret models with ill-defined factors, solutions containing five or more factors were all dismissed on conceptual grounds. Within Kinyarwanda, the three-factor solution consisted of factors labelled expectancy beliefs, combining mainly the expectancy items, value beliefs, combing the daily- and general-utility items, and costs beliefs, combining emotional-cost, perceived-useless, and outside effort items. The corresponding mathematics model consisted of factors labelled motivation, combining expectancy and intrinsic-value items, effort, combining mainly effort-belief items, and a third factor (demotivation) bringing together especially items on perceived useless and some cost aspects. Within both Kinyarwanda and mathematics, the four-factor model was the one with the smallest number of items (i.e., six) with double loading. Across both domains, this solution had at least three items on each factor that all exhibited strong loadings (> 0.40). Also, especially within the Kinyarwanda domain, items loaded on factors as hypothesized (e.g., McCoach et al., 2014). Across both domains, the four-factor solution also exhibited the same structure: one factor that mainly consisted of expectancy-belief items, one cost factor combining both effort-belief and emotional-cost items, one value factor combining daily- and general-utility items, and a fourth factor on perceived useless consisting of corresponding items. The four-factor solution was therefore considered the conceptually most plausible and statistically most sound model across Kinyarwanda, and mathematics. Within mathematics, the expectancy-belief factor was weakly correlated with the cost factor ($\rho = -0.31$) and moderately correlated with the utility-value factor ($\rho = 0.5$). Within both the Kinyarwanda, it was somewhat correlated with utility value ($\rho = 0.44$). Besides, only cost beliefs and perceived uselessness were weakly associated ($\rho = 0.47, 0.33$).

Table 13: Goodness of model fit (scale testing phase; EFA study I)

# factors	Kinyarwanda				Mathematics			
	RMSEA	CFI	TFI	SRMR	RMSEA	CFI	TFI	SRMR
3	.044	.963	.955	.073	.067	.93	.916	.079
4	.031	.983	.978	.057	.050	.962	.952	.053
5	.026	.988	.984	.051	.044	.973	.963	.046
6	.024	.991	.987	.046	.038	.981	.972	.041
7	.02	.994	.991	.041	.035	.985	.976	.036

Discussion

Even though dimensionality assessment and fit indices did not converge on the same number of factors (seven versus three and four), the four-factor model in particular appeared to be tenable from a conceptual point of view. The four factors of expectancy, costs, utility value, and perceived uselessness appeared to be well defined and described the data well. In this respect, the emerging latent structure somewhat resembled the expectancy-value-cost (EVC) model that has been increasingly reported in the literature (Barron & Hulleman, 2015; Flake et al., 2015; Kosovich et al., 2015). It resembled it in the sense that both the four-factor model and the EVC model contained a cost element in addition to the traditional components of expectancy and value. Within the four-factor model, the cost element seemed to be broken down into an effort belief/ emotional cost factor as well as a perceived uselessness factor (see table 14). Both factors exhibited weak correlations within both the Kinyarwanda and mathematics domain. Besides, the multi-faceted conceptualisation of cost beliefs (Flake et al., 2015) discussed in the literature did not materialize within Rwanda. For example, the cost facet pertaining to outside-effort costs did not emerge at all.

Table 14: Factor loadings within final model (scale testing phase; EFA study I)

Kinyarwanda		Ex	Co	Va	PU
Are you capable of [S]?	KAB6	0.93	0.05	-0.03	-0.10
Do you understand [S]?	KAB2	0.81	-0.05	0.07	0.01
Do you know the [S]?	KAB4	0.79	0.08	0.05	-0.10
Do you understand [S]?	KAB8	0.72	-0.19	0.04	0.15
Do you get good marks in [S]?	KAB5	0.71	-0.02	0.06	-0.07
Do you do well in [S]?	KAB3	0.70	0.03	0.15	-0.11
Is the subject of [S] easy?	KEF4	0.45	-0.22	0.14	0.17
Is the subject of [S] hard?	KEF3	-0.15	0.86	0.01	-0.14
Is the subject of [S] difficult?	KEF2	-0.30	0.81	0.05	-0.10
Does this [S] discourage you?	KNE3	0.14	0.80	-0.08	0.18
Does this [S] make you sad?	KNE2	0.16	0.80	-0.01	0.21
Does this [S] make you scared?	KNE4	0.19	0.74	0.06	0.21
Is the subject of [S] tiring?	KEF5	-0.21	0.71	0.12	0.09
Are you too busy to study [S]?	KOE2	0.08	0.64	0.00	0.01
Do you hate this [S]?	KNE1	0.09	0.62	-0.19	0.31
Do you fail in [S]?	KAB7	-0.14	0.56	-0.15	0.18
Is there too much work at home to study [S]?	KOE1	0.09	0.43	0.13	0.19
Is [S] good for your future?	KGE3	0.00	0.11	0.93	-0.01
Will [S] help you in your future life?	KGE1	-0.01	-0.09	0.83	0.06
Will [S] help you in what you will be?	KGE4	0.06	0.15	0.79	-0.10
Will [S] help you to get a good future?	KGE5	0.02	0.07	0.79	-0.04
Does [S] help you in your daily life?	KDA1	-0.01	-0.04	0.70	0.09
Do you use [S] in your everyday life?	KDA2	-0.11	-0.19	0.64	0.16
Is [S] useful for your everyday life?	KDA3	-0.05	-0.06	0.64	-0.10
Does this [S] make you proud?	KPO4	0.23	-0.10	0.59	-0.05
Does this [S] make you feel good?	KPO5	0.28	-0.17	0.54	0.12
Is [S] useful for your future?	KGE2	0.11	-0.03	0.51	-0.02
Do you think [S] is useless for you?	KPI2	-0.08	0.03	-0.03	0.90
Do you think [S] has no use for you?	KPI3	-0.02	0.10	0.00	0.88
Do you think [S] is not getting you anywhere?	KPI4	-0.10	0.13	0.07	0.78

Mathematics		Ex	Va	Co	PU
Do you understand [S]?	MAB2	0.95	-0.06	0.03	-0.07
Do you get good marks in [S]?	MAB5	0.89	0.00	0.00	-0.04
Do you know the [S]?	MAB4	0.88	-0.05	-0.06	0.03
Do you do well in [S]?	MAB3	0.82	-0.05	-0.03	-0.01
Are you capable of [S]?	MAB6	0.82	0.12	0.05	-0.05
Do you understand [S]?	MAB8	0.67	0.14	-0.10	-0.08
Do you like/ love this [S]?	MPO2	0.64	0.23	-0.07	0.09
Do you find time to study on [S]?	MOE3	0.55	0.16	-0.01	-0.15
Is the subject of [S] easy?	MEF4	0.55	0.09	-0.19	0.08
Does this [SUBJECT] make you proud?	MPO5	0.48	0.25	-0.09	0.09
Is [S] good for your future?	MGE3	0.03	0.91	-0.08	0.04
Will [S] help you in your future life?	MGE1	0.08	0.82	0.14	-0.02
Is [S] useful for your everyday life?	MDA3	-0.09	0.82	-0.02	-0.04
Will [S] help you in what you will be?	MGE4	0.04	0.82	-0.07	0.05
Will [S] help you to get a good future?	MGE5	0.17	0.82	0.10	0.12
Does [S] help you in your daily life?	MDA1	-0.10	0.81	-0.07	-0.07
Do you use [S] in your everyday life?	MDA2	0.02	0.64	0.09	-0.06
Is [S] useful for your future?	MGE2	0.22	0.62	0.02	0.01
Is the subject of [S] difficult?	MEF2	-0.15	0.01	0.83	-0.16
Is the subject of [S] tiring?	MEF5	-0.04	0.11	0.82	-0.03
Is the subject of [S] hard?	MEF3	-0.07	0.06	0.76	-0.02
Do you fail in [S]?	MB7	-0.07	0.08	0.69	0.00
Does this [S] make you sad?	MNE2	0.13	-0.21	0.65	0.26
Does this [S] make you scared?	MNE4	0.06	-0.07	0.62	0.22
Does this [S] discourage you?	MNE3	0.04	-0.20	0.59	0.25
Is there too much work at home to study [S]?	MOE1	0.05	-0.06	0.33	0.24
Do you think [S] has no use for you?	MPI3	-0.07	0.06	-0.03	0.91
Do you think [S] is not getting you anywhere?	MPI4	-0.02	0.08	-0.02	0.91
Do you think [S] is useless for you?	MPI2	-0.06	-0.03	0.07	0.83
Do you think you don’t give any value to [S]?	MPI1	0.01	-0.13	0.05	0.72

Note. Ex = Expectancy beliefs; Va = Value beliefs; Co = Cost beliefs; PU = Perceived uselessness. Also, item KDA2 should have been included into the language value scale but was not due to an oversight. ‘Do you understand [subject]’ is included twice. This is due to translation of the Kinyarwanda words ‘gusobanukirwa’ (i.e., to understand/ to be explained to) and ‘kumva’ (i.e., to understand/ to listen/ to hear). Also see footnote 11.

Most remarkably, however, was that intrinsic value did not emerge as a separate factor. Within the motivational literature, value has been conceptualised as highly multi-dimensional (Gaspard et al., 2015). In this respect, intrinsic value constitutes a key component. By contrast, amongst Rwandan secondary students, value appeared as a distinct factor only in form of utility value. Across both Kinyarwanda and mathematics, items measuring intrinsic value exhibited substantial double loading across both expectancy and value beliefs. Within self-concept research competence self-perceptions were often evaluated not only through items related to one's cognitive self-evaluations but also through items about one's own affective reactions towards a given subject (e.g. Marsh, 1999; Marsh et al., 2002). In fact, even though more recent evidence suggest that both affect and cognitive beliefs are best treated as separate constructs (Pinxten et al., 2014), some researchers have pointed out it is not clear yet to what extent both factors are indeed distinct (Bong & Skaalvik, 2003). Further exploratory work is required to study this question within the Rwandan context as well. It is best performed using both exploratory and confirmatory factor analysis. Sample size in this respect is best increased substantially. In particular when data-collection tools are categorical in nature, sample sizes of less than 400 may just not suffice. This may also explain why the results of both parallel analysis and dimensionality assessment based on fit indices have not converged.

5.2.3.3. Study 2: Exploratory factor analysis of revised school subject-specific motivational scales (EFA II)

As part of the project on achievement motivation in Rwanda (also see manuscript two for some more information), cross-sectional data on motivational beliefs across Chemistry, English, Kinyarwanda, and mathematics from 1378 lower-secondary students were collected. For 819 students, school marks were available as well. This subset was used to test the applicability of the GI/ E model within the Rwandan context (see manuscript two). The remaining subset with 559 students was then used to perform a second exploratory factor analysis (EFA) of the motivational instruments that emerged during the scale-development process outlined in S1 through S3. Whereas the previous EFA (see section S3) explored factor structures of scales within Kinyarwanda and mathematics, the EFA reported below explored latent structures of scales within Chemistry, and mathematics (two quantitative domains) as well as English, and Kinyarwanda (two language domains). Given that the sample size was also substantially bigger, latent structures were studied not only separately within each of the four domains but also across two different domains. For that purpose, altogether six exploratory factor analyses were performed that each targeted two out of the four domains. Hence, results below provide additional evidence on the latent structure not only within a wider set of school domains but also on the factor structure across domains.

Method

Procedure, Participants, and Instruments

Section 2.1 of manuscript two provides details on the data collection process. It also provides details about project schools recruited. Across the four schools, 1378 lower-secondary students completed the questionnaire on domain-specific motivational beliefs. For 559 of those students (around 40.6%), no achievement data was available. They therefore formed the subset used for the exploratory factor analysis reported below. Since this subsample was not randomly selected, they are systematically different from the remaining 819 students on a number of different characteristics. In comparison, they were younger ($\mathrm{M} = 13.9$ years; $\mathrm{SD} = 1.50$), less likely to be female ($\mathrm{M} = 45.6\%$), less likely to be from a boarding school ($\mathrm{M} = 49.2\%$), and were mainly enrolled in S1 ($\mathrm{M} = 79.2\%$) rather than S2 ($\mathrm{M} = 18.6\%$) or S3 ($\mathrm{M} = 2.1\%$). Students were enrolled across 27 classes with an average number of 20.7 ($SD = 19.4$) students per class (range from one to 68). The four mathematics scales described in section S3 were employed not only measure motivational beliefs in mathematics but also in Chemistry. The four Kinyarwanda scales described in section S3 were used to measure motivational beliefs both in Kinyarwanda and English. Students therefore completed two sets of four scales with 25 items each and two sets of four scales with 23 items each. Hence, EFA was performed on 96 items across 16 scales in total. The answer format again included the following three answer options: 1) always; 2) sometimes; 3) never. Motivational data was eventually coded in such a way that higher numbers indicated higher motivational levels.

Exploratory Factor Analysis

The statistical analysis proceeded as described in section S3. Hence, descriptive statistics for all study items were compiled first. Item correlations were compiled using polychoric correlation. Around 2.2% of motivational data from 559 students was missing. Missing values were imputed using chained equations in R (i.e., the mice package; van Buuren & Groothuis-Oudshoorn, 2011). Polychoric correlation analysis was then performed on each of the imputed datasets using the *polychoric* function in the *psych* package (Revelle, 2020) in R 3.6.2 (The R Core Team, 2019). The individual correlation matrices were then pooled. Given a data set of 559 secondary students, the sample was not big enough to simultaneously explore the latent structure of 16 scales with 96 items across four school domains. Instead, the data set of 559 students was further divided into six subsets (each with 559 observations) that each combined motivational data from two different domains (see table 15). Parallel analysis was performed for each of the six combinations based on the pooled correlation matrix. Parallel analysis was implemented using *fa.parallel* of the psych package (Revelle, 2020) using the 95th per centile and principal axis factoring (e.g., Finch, 2020; Garrido et al., 2013). EFA was largely performed in the way outlined in section S3.2.2

using Mplus 8.7 (e.g., DWLS; L. K. Muthén & Muthén, 2017). Since students were now nested within 27 classes, factor analyses were performed this time with a design-based correction of standard errors (i.e., type = complex; e.g., Ji et al., 2021). To deal with these missing data, Mplus implemented the FIML approach, in which all available information is taken into account when the model parameters are estimated (Asparouhov & Muthén, 2010). Models were evaluated using the fit indices RMSEA, CFI, and TFI in particular (Finch, 2020).

Table 15: Results of parallel analysis (scale testing phase; EFA study II)

# EFA	Domains Combined		Missing Values	Results Parallel Analysis
1	Chemistry	English	2.3%	7
2	Chemistry	Kinyarwanda	1.9%	7
3	Chemistry	Mathematics	2.6%	6
4	English	Kinyarwanda	1.9%	7
5	English	Mathematics	2.6%	7
6	Kinyarwanda	Mathematics	2.2%	7

Results

Descriptive Statistics

Across the scales, means for each item were high with 2.36 ($SD = 0.23$) within Chemistry, 2.39 ($SD = 0.29$) within mathematics, 2.56 ($SD = 0.18$) within English, and 2.61 ($SD = 0.17$) within Kinyarwanda. Hence, at the item level students appeared to be more motivated within languages than within the quantitative domain. Again, means were particularly high amongst value ($\mathrm{M's} = 2.72;\ \mathrm{SD} = 0.12$) compared to expectancy ($\mathrm{M's} = 2.45;\ \mathrm{SD} = 0.2$), cost ($\mathrm{M's} = 2.37;\ \mathrm{SD} = 0.23$), and perceived uselessness ($\mathrm{M's} = 2.24;\ \mathrm{SD} = 0.1$). Value items also tended to be more (negatively) skewed ($\mathrm{M's} = -1.82;\ \mathrm{SD} = 0.57$) and to exhibit stronger kurtosis ($\mathrm{M's} = 5.83;\ \mathrm{SD} = 2.3$) compared to expectancy ($M = -.62\ (0.53);\ M = 2.82\ (0.59)$), costs ($M = -0.75\ (0.62);\ M = 2.95\ (0.93)$), and perceived uselessness ($M = -0.5\ (0.21);\ M = 1.59\ (0.19)$). Factor analysis based on DWLS estimation, however, is robust to non-normal data (Finney & DiStefano, 2013). Average intra-cluster correlation (ICC) was .006 and ranged from 0 to 0.2. Table 16 presents a summary of the correlations within each provisional scale. Across all scales, the average correlation was above 0.60 and ranged from 0.33 to 0.86. It appeared to be particularly high within perceived uselessness with 0.72. Average intra-scale correlations were also above 0.60 across the different domains and were particularly high in English with 0.72. Average correlations seemed to be somewhat higher within the language (0.69) than the quantitative domain (0.62).

Table 16: Summary of the intra-scale correlations (scale testing phase; EFA study II)

Subject	Range	Expectancy	Utility value	Uselessness	Costs
Chemistry	Mean	0.60	0.61	0.70	0.58
	Min	0.45	0.44	0.62	0.46
	Max	0.76	0.76	0.78	0.70
English	Mean	0.78	0.53	0.86	0.71
	Min	0.70	0.33	0.84	0.65
	Max	0.86	0.33	0.87	0.77
Kinyar-wanda	Mean	0.59	0.77	0.59	0.70
	Min	0.46	0.71	0.37	0.50
	Max	0.72	0.81	0.76	0.77
Mathematics	Mean	0.59	0.59	0.72	0.55
	Min	0.35	0.35	0.67	0.40
	Max	0.78	0.78	0.83	0.74

Note. Mean was calculated based on absolute values of all intra-item correlations

Exploratory Factor Analysis

As table 15 on page 162 highlights, across the different combinations parallel analysis almost consistently suggested seven factors across the eight scales. For each of the six domain combinations, models with three through eight factors were estimated. Fit indices are presented in table 17 on the next page. Most models exhibited good to excellent fit. Applying

RMSEA 0.15 difference value to identify models (Finch, 2020) did not suggest any model with more than three factors across the four domains. Hence, dimensionality assessments through parallel analysis and fit indices again did not converge in terms of their results (i.e., seven versus three or less factors).

Table 17: *Goodness of fit (scale testing phase; EFA study II)*

Subject combination	Factor solution	Fit indices			
		RSMEA	CFI	TFI	SRMR
Chemistry/ English	3F	0.037 (0.034 – 0.04)	0.911	0.898	0.095
	4F	0.028 (0.025 – 0.031)	0.950	0.941	0.074
	5F	0.019 (0.015 – 0.023)	0.978	0.972	0.049
	6F	0.016 (0.01 – 0.02)	0.986	0.982	0.040
	7F	0.013 (0.005 – 0.019)	0.991	0.987	0.036
	8F	0.011 (0 – 0.017)	0.994	0.991	0.032
Chemistry/ Kinyarwanda	3F	0.039 (0.036 – 0.042)	0.879	0.862	0.103
	4F	0.03 (0.027 – 0.033)	0.931	0.918	0.077
	5F	0.023 (0.019 – 0.027)	0.961	0.951	0.060
	6F	0.015 (0.009 – 0.02)	0.984	0.979	0.040
	7F	0.012 (0.001 – 0.018)	0.990	0.987	0.035
	8F	0.01 (0 – 0.016)	0.994	0.991	0.032
Chemistry/ Mathematics	3F	0.04 (0.037 – 0.043)	0.877	0.860	0.106
	4F	0.028 (0.025 – 0.031)	0.941	0.930	0.070
	5F	0.02 (0.015 – 0.023)	0.973	0.966	0.048
	6F	0.015 (0.009 – 0.02)	0.984	0.980	0.039
	7F	0.014 (0.006 – 0.019)	0.988	0.984	0.036
	8F	0.012 (0.002 – 0.018)	0.991	0.987	0.032
English/ Kinyarwanda	3F	0.039 (0.036 – 0.042)	0.911	0.898	0.105
	4F	0.028 (0.025 – 0.032)	0.957	0.948	0.076
	5F	0.021 (0.017 – 0.025)	0.977	0.971	0.055
	6F	Did not converge			
	7F	0.013 (0.001 – 0.018)	0.993	0.990	0.036
	8F	0.011 (0 – 0.017)	0.995	0.993	0.033
English/ Mathematics	3F	0.038 (0.036 – 0.041)	0.907	0.894	0.106
	4F	0.03 (0.026 – 0.033)	0.947	0.937	0.073
	5F	0.021 (0.017 – 0.025)	0.975	0.969	0.056
	6F	0.016 (0.01 – 0.021)	0.986	0.982	0.042
	7F	0.013 (0.004 – 0.018)	0.992	0.988	0.036
	8F	0.011 (0 – 0.017)	0.994	0.991	0.034
Kinyarwanda/ Mathematics	3F	0.028 (0.025 – 0.031)	0.877	0.859	0.113
	4F	0.031 (0.028 – 0.034)	0.933	0.920	0.085
	5F	0.025 (0.021 – 0.029)	0.959	0.949	0.065
	6F	0.016 (0.011 – 0.021)	0.983	0.978	0.043
	7F	0.014 (0.006 – 0.019)	0.989	0.985	0.037
	8F	0.011 (0 – 0.017)	0.993	0.990	0.034

A conceptual interpretation of the different models based on the simple-structure criterion (Brown, 2015, p. 27) came to rather consistent results across the six domain combinations. Within seven-factor models, each of the factors were well defined in that items hardly exhibited any cross-loadings. Also, within each scale all items loaded on the same factor and only that factor. This suggested that the empirical internal structure associated with the motivational scales met their theoretical conceptualisation. The only exception in this respect were the items pertaining to perceived uselessness. Whatever the domain combination, all items on perceived uselessness from both subjects loaded on the same factor. Between the six subsets, the average loading of perceived-uselessness items on the cross-domain factor was 0.83. The domain-specific expectancy, utility value, and cost factors exhibited average loadings in the range from 0.52 to 0.85. Hence, with the sole exception of perceived uselessness all other scales were well defined across the different domains.

Table 18: Latent factor correlations (scale testing phase; EFA study II)

	1 C-E	2 C-V	3 C-C	4 E-E	5 E-V	6 E-C	7 K-E	8 K-V	9 K-C	10 M-E	11 M-V	12 M-C	13 Use
1													
2	**0.43**												
3	0.25	0.25											
4	0.14	**0.31**	0.25										
5	0.08	0.30	0.27	0.24									
6	0.02	0.03	0.22	0.20	0.11								
7	0.24	0.15	0.14	0.22	0.10	0.07							
8	0.11	0.25	0.10	0.10	0.26	-0.02	**0.33**						
9	0.02	0.09	0.24	0.10	0.09	0.20	**0.34**	0.20					
10	**0.33**	0.12	0.06	0.13	0.06	0.04	0.22	0.09	0.10				
11	0.16	**0.33**	0.09	0.13	0.29	0.13	0.19	0.24	0.16	**0.36**			
12	0.09	0.09	0.26	0.02	-0.04	0.16	0.07	0.11	0.27	**0.42**	0.23		
13	0.06	0.16	0.24	**0.33**	0.11	0.27	0.03	0.12	**0.37**	0.22	0.15	0.23	

Note. Within column names, first letter refers to school domain (e.g., C= Chemistry) whereas the second one refers to scale (e.g., C = cost beliefs). Use refers to perceived uselessness.

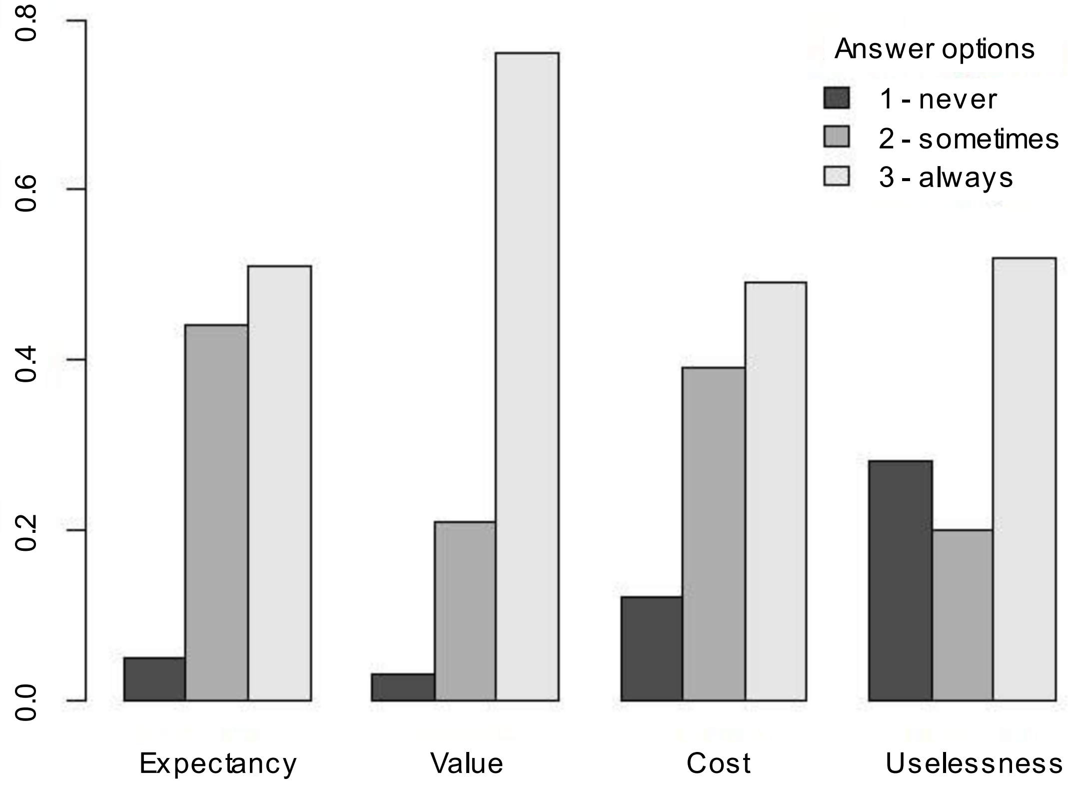
Table 18 presents the latent-factor correlations extracted from EFA's of the six domain combinations. For example, correlations between English- and mathematics-related factors were extracted from the EFA of the English/ mathematics motivational subset. Each domain-specific factor, however, was correlated with the domain-general factor perceived uselessness three times across the six domain combinations. Thus, correlations between perceived uselessness and other motivational scales represent averages. Within domains, expectancy tend to correlate with value as well as cost beliefs. Within English, factors did not appear to be correlated. Across Chemistry and mathematics (both quantitative domains), expectancy and value were each correlated with the corresponding factor within the other domain. With the only exceptions of English expectancy beliefs and Kinyarwanda costs, uselessness appeared to be uncorrelated with other motivational factors.

Discussion

The second EFA of the provisional scales on achievement motivation amongst Rwandan lower-secondary students was an extension of the previous in several respects. First, it was based on a substantially larger student sample (559 versus 371 students). Second, instruments that emerged from the scale-development process were evaluated not only within the initial domains of Kinyarwanda and mathematics but also within the two additional domains of Chemistry and English. Hence, the internal structures of the scales were now evaluated in two subjects within the language and quantitative domain. Third, factor evaluations of scales were not only performed within each school domain. Through a series of six separate EFAs, the latent structure of scales was always evaluated jointly across two different subjects. Fourth, the sample consisted of students from 27 classes altogether. 27 was deemed sufficient to control for the nested data structure through design-based correction of standard errors (i.e., type = complex). Hence, the second EFA may constitute a more robust evaluation of the provisional scales than the first EFA reported in section S3.

Especially factor models with more than four factors showed excellent model fit. Across the different domain combinations, parallel analysis almost always suggested seven factors to adequately summarize the latent structures of scales. This matched the theoretical interpretations of the different factor models that emerged. Across all domain combinations, the seven-factor model consistently exhibited a simple structure involving factors that were well-defined especially in the case of scales on expectancy, utility-value, and cost beliefs. Corresponding items exhibited strong loadings on those factors associated with their scales. Also, cross loadings were only found in some cases and also tended to be small in case of the second factor. As expected, expectancy beliefs generally correlated with value as well as cost beliefs. Moreover, both expectancy and value beliefs were each correlated across subjects within the quantitative domain (i.e., Chemistry and mathematics). This is expected because both Chemistry and mathematics can be considered to be rather similar domains. By contrast, expectancy, value, and cost beliefs were not correlated across subjects within the language domain. This is not unexpected. Within the literature, it has been increasingly highlighted that first- and second-language subjects do not necessarily constitute subjects perceived to be similar (e.g., van der Westhuizen et al., 2022). Also, that motivational beliefs were not correlated between the language and quantitative domain is in line with the literature. It has been argued that both domains generally constitute rather different school subject areas, something discussed within the literature review of manuscript two. Accordingly, motivational beliefs in one area tend to be not predictive of motivational beliefs in the other area. Hence, the current analysis provides some evidence on the internal and external validity of the scores obtained through the expectancy, value, and cost scales.

Figure 12: Student answers within motivational scales (scale testing phase; EFA study II)



Results were more mixed within the case of perceived uselessness. Within each of the six domain combinations, scales pertaining to perceived uselessness loaded on a cross-domain factor only. Loadings exhibited in this respect also tend to be strong. This is rather unexpected, because motivational constructs are generally conceptualised to rather domain specific. Also, the cross-domain factors on perceived uselessness generally seemed to be uncorrelated with other motivational factors with the sole exception of English expectancy beliefs and Kinyarwanda cost beliefs. One explanation of these results concerned how these items interacted with the answer options. Like cost items, items on perceived uselessness were negatively worded. However, whereas cost items contained negative descriptors such as being sad or difficult, items on perceived uselessness are constructed using the negating word *ntabwo* (not/ never) or *nta* (i.e., there is no). For example, the third item reads 'Ese [SUBJECT] *nta*cyo [ri]kumariye?' (i.e., Does [SUBJECT] have no use for you?). Potentially, it was not clear to students how to use answer options provided (i.e., always/ sometimes/ never) to express their perceptions vis-à-vis the uselessness of a subject. Hence, for some students the answer option ‘always’ may have implied that a subject is indeed use*less* whereas for others it may have stood for a subject being use*ful*. Figure 12 above may illustrate this. It depicts the average probabilities of students any of the three answer options within the different scales. Within expectancy, value, and cost scales, the more positive the answer option the more likely it was chosen. Conversely, within perceived uselessness, both ‘1’ and ‘3’ were both more common than the middle ground ‘2’. Against this background, the cross-domain factors associated with the perceived-uselessness scale could be interpreted as some form of response style amongst students. It was unfortunate that student feedback during cognitive interviews (see section S2) did not catch this potential cause of confusion. One reason could have been that students were more attentive to the type of questions asked and thus were more likely to guess what the questions on perceived uselessness were about and how to answer them best. Given that perceived uselessness has been increasingly recognized as a driver of school dropouts (e.g., Ailei, 2019), additional research on its potential effects on student disengagement certainly appears warranted. However, its combination with the answer options of always; sometimes; never does not seem to have worked. They prevent a consistent interpretation of the data. Additional research on how to best phrase the answer options is necessary.

Conclusion

Evidence based on a second-round EFA suggests that the motivational scales on expectancy, value and cost beliefs adequately measure respective constructs amongst Rwandan lower-secondary students. By contrast, the items on perceived uselessness seemed to be incompatible with the answer options provided. This in turn may have caused confusion amongst students on how to best answer the questions on perceived uselessness using the three answer options provided. Hence, rather than representing any motivational constructs the factor associated with scales on perceived useless may be best interpreted as some form of method factor that expresses how students deal with this incompatibility between items and answer options. It was therefore decided to exclude all data on perceived uselessness from the main analysis reported in manuscript two. Hence, the main analysis focussed only on expectancy, utility value, and cost beliefs across the different school domains. The final scales used are presented in annex 3.

5.2.3.4. Study 3: Exploratory factor analysis of lesson-specific motivational scales (EFA III)

The purpose of the third EFA study of the scale-testing phase was to explore the factor structure (i.e., structural validity) of the lesson-specific motivational self-report items. To do so, student self-reports were collected at a basic-education school outside Kigali. Unlike in the previous two EFA studies, data collection was not limited to just two, and four subjects respectively. Thus, students provided self-reports across almost all subjects studied at the secondary level. This study largely followed the procedure (i.e., single-level exploratory factor analysis) of the previous EFA studies. As the results below highlight, the lesson-specific scales slightly differed in terms of the motivational themes covered as well as in terms of the items they included.

Method

Procedure, Participants, and Instruments

Data for study two was collected during two days in the middle of the first week of teaching of the 2020 school year. For this study, a public school was randomly selected out of the school register of one of the districts outside Kigali. The school selected is located in a semi-rural part of the district. At that time, students in the first grade (S1) of lower secondary have not started schooling yet. Thus, only S2 and S3 students were included into the study. Across the four S2 and S3 classes, 226 students were enrolled. To recruit students, the same procedure as outlined for study one was followed. 207 students (91.6 per cent) submitted valid consent forms and thus were included into the study. Across the two days of data collection, 197 eligible students (95.7 per cent) submitted questionnaires. Around 53 per cent of them were female with mean age of 15.4 years.

For data collection, lessons were randomly selected out of the school timetable for target classes. Hence, all school subjects except physical education could have been selected. According to the initial data-collection plan, each class would have been subject to 4.5 measurement occasions across the two days on average. Due to some teachers being absent from class unexpectedly, however, the mean measurement occasion was around three. As a result of the random lesson selection, all but one classes were subject to repeated data collection within the same subject. During lesson-specific data collection, questionnaire administration was handled by three survey teams each consisting of two enumerators that were hired from the district of the school. Questionnaires were only handed out to students once a teacher indicated to have closed the lesson (Ahmed, Minnaert, et al., 2010; Vasalampi et al., 2016). On each occasion, two enumerators entered the class and handed the questionnaires. To save time, no comments were made but that students always had to choose one of the answer options and that data was kept confidential.

The questionnaire on motivational states administered in study two consisted of seven scales and 36 items in total. The expectancy sub-scale consists of seven items (e.g. Do you feel you understood today's activities of [SUBJECT]?), the daily-utility sub-scale consists of three items (e.g. Do you think today's activities of [SUBJECT] will help you in your daily life?), the school-utility sub-section consists of six items (e.g. Do you feel today's activities helped you succeed in [SUBJECT]), the unimportance sub-scale consists of five items (e.g. Do you think today's activities of [subject] had no use for you?), the intrinsic-value sub-scale consists of five items (e.g. Did you like/ love today's activities of [SUBJECT]?), the effort sub-scale consists of five items (e.g. Were today's activities of [SUBJECT] difficult?), and the emotional-cost sub-scale consists of four items (e.g. Did you hate today's activities of [SUBJECT]?). Across both studies, questionnaire items were presented in random order. Based on feedback gathered during the qualitative interviews during phase one (see section 5.2.3.1 for more details), the item response format for both belief and state scales was a Likert scale with three options: 1) Most of the times; 2) Sometimes; 3) Never. Questionnaires were only given out in Kinyarwanda. Questionnaire administration took around 36.7 minutes per class during study one and around 12 minutes during study two.

Exploratory Factor Analysis

The (single-level) exploratory factor evaluation of the scale data was completed through the following process (e.g., McCoach et al., 2014; Preacher & MacCallum, 2003; Watkins, 2018; also see discussions of EFA studies 1 and 2). First, for each scale item means, standard errors, and inter-item correlations were determined. Low correlations of |0.3| were considered to justify removal of corresponding items (McCoach et al., 2014; Watkins, 2018).

Given item’s three-point response format, factor evaluations were based on the polychoric correlations (Olsson, 1979). Second, the number of latent factors to be extracted was determined (Fabrigar & Wegener, 2012; Gaskin & Happell, 2014; Osborne, 2014; Schmitt, 2011). The observed polychoric correlation matrix was analysed using scree plots, parallel analysis, and the minimum average partial procedure (McCoach et al., 2014). Third, the individual factor solutions were estimated and then compared in terms of their statistical fit as well as conceptual plausibility (Fabrigar & Wegener, 2012; Watkins, 2018). To account for the nonnormal, categorical data, all factor models were extracted from the polychoric correlation matrix using diagonally weighted least squares (DWLS) estimation (Finney & DiStefano, 2013). Since latent factors were expected to be partly correlated (Watkins, 2018), factor extraction was implemented with oblimin (i.e., oblique) factor rotation (Fabrigar & Wegener, 2012). Fit indices used included the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). Both the CFI and TLI are goodness-of-fit indices. Values of .95 or higher are seen as indicator of an excellent fit (West et al., 2012). By contrast, RMSEA is badness-of-fit indices. A RMSEA value of .06, or lower, is seen as evidence of indicative of good model fit (Hu & Bentler, 1999). However, especially overfactored models tend to exceed the aforementioned thresholds as well (D. A. Clark & Bowles, 2018). Incremental improvements within fit indices (“fit statistic difference approach”) between models were estimated as well (Finch, 2020). The conceptual plausibility of factor solutions was assessed through evaluating the pattern matrix (J. M. Graham et al., 2003; Watkins, 2018) using the “simple structure” (Brown, 2015, p. 27) criterion.

The data collected was hierarchically structured (e.g., Raudenbush & Bryk, 2002). A comment on adequacy of a single-level exploratory factor analysis of the data is thus warranted. To account for the dependency within the data, lesson-specific scales would have ideally been evaluated using multilevel exploratory factor analysis (D’Haenens et al., 2010;

Kim et al., 2016). However, a general challenge associated with MEFA is sample size. Whereas a student sample of 100 individuals is deemed sufficient, effective estimation would require at least five lesson-specific observations per student (Holtmann et al., 2016; Hox & Maas, 2001; Tobias Koch et al., 2014). Within the Rwandan context this would have implied data collection for at least a week. Due to budget and time constraints, however, this was not feasible. It was for this reason that lessons across all subjects were sampled. This may also have removed dependency in the data at least partly. In fact, whereas the number of questionnaires submitted per student was 2.45 questionnaires on average it was only about 1.12 on average per student and subject. During the main phase, lesson-specific self-reports were analysed through multilevel models (i.e., MLCFA; Kim et al., 2016; Rabe-Hesketh et al., 2012; Stapleton et al., 2016).

Table 19: Results of dimensional assessment (scale testing phase; EFA study III)

Method	Proposed number of factors
Scree Plot	6
Horn’s Parallel Analysis with 95% Per centile	7
UCP Adjusted Parallel Analysis	3
Minimum Average Partial Procedure	5

Results and concluding comments

The dimensional analysis of the raw data using the different eigenvalue-based methods suggested factor solutions with three to five factors (see table 19). A comparison of the fit indices of these models is presented in table 20. Especially the solutions with four and more factors generally exhibited good fit. The final model was therefore selected based on the fit statistic difference approach. Even though the RMSEA fit index did not improve substantially across models, there were significant improvements especially within the CFI and SRMR fit indices. Against this background, the five-factor model was chosen. Table 21 presents the factor loadings within the pattern correlation matrix associated with the final model. The model was also plausible from a conceptual point of view. The models that emerged were labelled as “daily utility value”, “effort”, competence”, “perceived unimportance”, and “school utility”. It was therefore decided to adopt four lesson-specific motivational scales. For each scale, the first three items associated with the corresponding factor were included. Thus, the final questionnaire to gauge the different dimensions of lesson-specific achievement motivation amongst Rwandan secondary students consisted of 15 items. However, reflections upon the scale on perceived uselessness discussed in the EFA study 2 above eventually led to the removal of the corresponding data. Hence, the analysis of the intraindividual data of the project’s main phase did not include lesson-specific perceived uselessness.

Table 20: Goodness of fit (scale testing phase; EFA study III)

Model	Fit indices			
	RMSEA	CFI	TFI	SRMR
3- Factor Model	0.058	0.932	0.918	0.068
4- Factor Model	0.047	0.957	0.945	0.056
5- Factor Model	0.041	0.970	0.958	0.046
6- Factor Model	0.037	0.977	0.966	0.040

Table 21: Factor loadings within final model (scale testing phase; EFA study III)

Items	Daily utility Value	Effort	Comp-etence	Perceived uselessness	School utility
Do you think today’s activities of [SUBJECT] will help you in your daily life?***	0.85	-0.05	-0.05	-0.01	0.03
Do you think you will use what you learned in today’s activities in [subject] in your everyday life? ***	0.85	0.00	-0.02	-0.05	0.02
Do you think today’s activities of [subject] will be useful for your everyday life? ***	0.65	-0.01	0.00	-0.05	0.18
Did today’s activities of [SUBJECT] make you happy?	0.53	-0.18	0.27	-0.05	0.11
Did today’s lesson of [SUBJECT] make you feel good?	0.50	-0.09	0.24	-0.06	0.15
Did you like/ love today’s activities of [SUBJECT]?	0.44	0.06	0.30	-0.24	0.19
Were today’s activities of [SUBJECT] hard? ***	-0.06	0.89	0.01	-0.03	0.02
Were today’s activities of [SUBJECT] tiring? ***	0.01	0.88	-0.06	-0.03	0.01
Were today’s activities of [SUBJECT] confusing? ***	-0.12	0.73	-0.01	0.05	0.10
Did today’s activities of [SUBJECT] make you have fear?	0.04	0.71	0.10	0.08	-0.04
Were today’s activities of [SUBJECT] difficult?	0.07	0.54	-0.20	0.22	0.15
Do you feel you understood today’s activities of [SUBJECT]? ***	0.00	0.13	0.84	-0.22	0.07
Do you feel you were capable to do today’s activities of [SUBJECT]? ***	-0.12	-0.05	0.78	0.04	0.09
Do you feel you understood today’s activities of [SUBJECT]? ***	0.07	-0.08	0.74	0.01	0.05
Do you feel you succeeded in today’s activities of [SUBJECT]?	0.06	-0.06	0.67	0.16	0.11
Do you feel you did well in the questions of today’s activities of [SUBJECT]?	0.17	-0.09	0.62	0.23	0.12
Do you feel you knew today’s activities of [SUBJECT]?	0.23	-0.13	0.52	0.03	0.20
Were today’s activities of [SUBJECT] easy?	0.15	-0.24	0.49	0.27	-0.06
Do you think today’s activities of [subject] were useless? ***	-0.17	-0.04	0.06	0.91	0.08
Do you think today’s activities of [subject] had no use for you? ***	-0.03	0.02	-0.08	0.86	0.05
Do you think you did not give any value to today’s activities of [subject]? ***	-0.18	0.10	0.09	0.79	-0.09
Do you think you did not care about today’s activities of [subject]?	0.18	0.04	-0.09	0.76	-0.03
Did you hate today’s activities of [SUBJECT]?	-0.02	0.28	-0.08	0.58	0.00
Do you think today’s activities of [subject] will not lead you to anywhere?	-0.00	0.24	0.12	0.51	-0.23
Do you feel today’s activities helped you prepare for the [SUBJECT] exams? ***	-0.03	0.01	-0.04	-0.01	0.94
Do you feel today’s activities helped you to do the exam of [SUBJECT]? ***	-0.02	-0.05	0.01	-0.01	0.91
Do you feel today’s activities helped you succeed in [SUBJECT]? ***	0.02	0.04	0.11	0.02	0.75
Do you feel today’s activities helped you to see if you can do well in [SUBJECT]?	0.14	0.09	0.09	0.01	0.69
Do you feel today’s activities helped you to understand [SUBJECT]?	0.18	0.04	0.18	0.03	0.53
Do you feel today’s activities helped you to know where you stand in [SUBJECT]?	0.30	-0.01	0.08	0.04	0.51
Do you feel failed in today’s activities of [SUBJECT]?	0.13	0.27	-0.43	0.30	0.19
Did today’s activities of [SUBJECT] make you sad?	0.26	0.31	-0.06	0.61	-0.19
Did today’s activities of [SUBJECT] make you be discouraged?	0.01	0.63	0.04	0.31	-0.22
Did today’s activities of [SUBJECT] make you happy?	0.38	0.02	0.48	-0.23	0.05
Did today’s activities of [SUBJECT] make you proud?	0.52	0.02	0.31	-0.01	0.06

Note. Model fit was RMSEA = 0.041; CFI = 0.975; TLI = 0.963; SRMR = 0.042. ‘***’ indicates that these items were included into scales.

5.3. Manuscript two: The Generalized Internal/External Frame of Reference Model among Secondary School Students in Rwanda.

Abstract

According to generalized internal/ external (GI/E) frame-of-reference model, motivational beliefs are explained through academic achievement. In Africa, respective studies are rare. In the present study, we investigated if the model explains expectancy, utility, and cost beliefs of Rwandan lower-secondary students (N = 769; 51.0% female) within Chemistry and mathematics (quantitative domain) as well as English and Kinyarwanda (language domain). Through multiple-group structural equation models (SEM) we compared the model's applicability to basic-education and boarding schools, which can be taken as two different ability-based tracks within Rwanda's education system. High performance on national school examinations qualifies for boarding schools whereas basic-education schools do not require any performance levels. The model's applicability differed across school types. Within basic-education schools, achievement predicted mainly cost beliefs. Within boarding schools, achievement predicted cost and expectancy beliefs in particular. Across both types, respective beliefs were positively predicted by achievement within subjects. Within basic-education schools, beliefs within one language were also positively predicted by achievement in the other language (i.e., assimilation effects). Within boarding schools, beliefs within subjects of one of the domains (i.e., language or quantitative) were negatively predicted by prior achievement in subjects of the other domain (i.e., contrast effects). We therefore concluded that school-level contextual factors such as the salience of the English language within classrooms may moderate motivational processes that Rwandan secondary students experience. This may have implications especially for the design of motivational interventions whose potential has not been fully explored within the African context.

Keywords: Internal/external frame of reference model, expectancy-value theory, Africa, secondary education, school-level ability tracking

Introduction

Students' motivational beliefs in form of expectancy, value, and cost perceptions have been found to predict academic choices and learning behaviours within school contexts (e.g., Eccles & Wigfield, 2020). It is therefore important to understand their antecedents. According to the generalized internal/external frame of reference (GI/ E) model, prior academic achievement shapes motivational beliefs within and between school subjects. Within subjects, academic achievement is assumed to positively predict motivational beliefs. Between subjects, academic achievement in one subject is assumed to predict motivational beliefs in the other subject negatively (i.e., contrast effects) if the respective subjects are perceived to be different (e.g., mathematics versus English). To the extent that the two subjects are considered similar (e.g., mathematics and Physics), respective effects can be positive (i.e., assimilation effects; Möller et al., 2016). The applicability of the GI/ E model especially to expectancy beliefs (i.e., often operationalised through academic self-concept) has been demonstrated in many parts of the world (Marsh & Hau, 2004; Möller et al., 2009, 2020). A major gap in this respect appears to be that Africa remains largely underrepresented within GI/E model research (Marsh & Hau, 2004). The only investigation that empirically tested the model within the African context might have been the multi-country study reported by Chiu (2008). Using TIMSS data, she examined relations between achievement and expectancies (self-concept) between mathematics and Science also within the contexts of Botswana, Ghana, and South Africa. She found country-specific variations. In Ghana, cross-domain effects were positive whereas in Botswana and South Africa they were negative, which suggests that country contexts may moderate the relations between prior achievement and motivational beliefs at least to some extent. This in turn implies the question about how GI/ E model relations are born out in other African contexts. In the present study, we applied the GI/E model to motivational beliefs reported by lower-secondary students from Rwanda, East Africa. We extended Chiu's (2008) research in four ways. First, we did not only include two subjects from both the scientific domain (Chemistry and mathematics) but also English and Kinyarwanda, the former being Rwanda's main medium of school instruction and the latter being Rwanda's national language. Second, we did not only include expectancy beliefs but also utility perceptions, and cost beliefs. We were therefore one of the first studies to employ an expectancy-value framework (Eccles & Wigfield, 2020) to study student motivation within an African context (also see Bofah & Hannula, 2015). Third, we controlled for school type (i.e., basic education versus boarding schools). In Rwanda, there is a de-facto school-level ability grouping of students (e.g., Trautwein, Lüdtke, Marsh, et al., 2006). Boarding schools are considered post-primary schools for high-achieving students that performed well in national examinations whereas basic-education schools are considered to be for those who "failed" (Williams, 2019, p. 646). Our study provides insights into the extent to which school type moderates GI/E model relations. Given that a de-facto ability grouping is common within many African education systems (Sayeda & Kanjeeb, 2013), this study may therefore be of relevance to other countries of the continent as well. Fourth, we relied on locally grounded scales to gauge student motivation in a cultural-linguistically sensitive way. Employing Western-borne instruments in non-Western contexts has often been associated with cultural bias and low reliabilities (Bofah & Hannula, 2015; Marsh et al., 2013; Zusho & Clayton, 2011), caveats that at least partly emerged in Chiu's (2008) study. Hence, the present study can also be taken as a first attempt to measure student motivation within Rwanda quantitatively.

Motivational Beliefs within Expectancy-Value Theory

According to Eccles' et al. (e.g., 2020) expectancy-value theory, students' academic choices, and learning engagement is predicted by both their expectancy and value perceptions. Expectancies refer to those self-perceptions about how well one is going to perform academically. They are assumed to be shaped by beliefs about one's own competencies in a given domain (i.e., "self-concept of ability"; Eccles et al., 1983; Wigfield & Eccles, 2000). Self-concepts in turn are the product of self-evaluation of one's own past performance in that domain (Eccles, 2009). Empirically, both expectancy and ability self-concept beliefs tend to be inseparable factors (e.g., Eccles, Wigfield, et al., 1993) and thus we use them interchangeably in the present study (e.g., Wigfield & Gladstone, 2019). Value perceptions are about the perceived level of enjoyment an academic challenge brings about (i.e., "intrinsic value"), how useful learning outcomes associated with that challenge are perceived to be for oneself (i.e., "utility value"), and how important successful completion of the challenge is for one's own self-image (i.e., "attainment value"; Eccles et al., 1983). Unlike intrinsic and attainment value, utility perceptions are considered to be malleable. For example, evaluation studies have shown that they can be shifted through targeted interventions with positive knock-on effects on achievement (Harackiewicz et al., 2016; e.g., Hulleman, Godes, et al., 2010). In more recent years, it has also been increasingly discussed whether or not academic choices and learning behaviours are shaped by students' perceptions about how cognitively and emotionally taxing these choices and behaviours are (i.e., cost beliefs; Barron & Hulleman, 2015; Flake et al., 2015). The implicit assumption is that the more effortful and emotionally taxing a challenge is the more likely it is that students abstain from it.

Motivational beliefs are considered to be not only multidimensional in the way just outlined but also school-subject specific.[19] Subject specificity implies that motivational beliefs are only weakly or even negatively correlated between those school subjects that are considered to be different (Gaspard et al., 2018; Trautwein et al., 2012). Especially within early self-concept research, near-zero correlations between competence beliefs of different subjects were deemed "paradoxical" (Marsh, 1990, p. 139; Marsh & Hau, 2004). They were initially deemed paradoxical because academic self-concept tends to be strongly related with academic achievement while achievement is generally strongly correlated across school subjects (Möller et al., 2009, 2020). To explain these paradoxical relations (i.e., school subject-specificity) between self-concepts associated with different subjects, the internal/external frame of reference (I/ E) model was proposed (Marsh, 1986; Marsh et al., 1988).

The GI/ E Model of Motivational Beliefs

The GI/ E model is an extension of the classic I/ E model, which explains the levels of self-concept exhibited within the mathematics and language domain. It hypothesizes that self-concept in each domain is based on self-evaluations carried out using the following two comparison points. First, students evaluate their own academic achievement in any of the two domains by contrasting it with achievements attained by significant others (e.g., classmates) in that domain (i.e., external references). Within external (i.e., social) comparison processes,

[19] Within the motivational literature, subject specificity of constructs is often referred to by the term of domain specificity (Möller et al., 2016). Within the present study, however, we understand a school domain (e.g., language domain) as a content area that may comprise of not only one but several different school subjects (e.g., English and Kinyarwanda). Domain-specificity may imply that correlations between subjects of the same domain are substantial. As the example of low correlations (ρ_{GE} = .18; $p < .05$) between German and English expectancy beliefs (i.e. academic self-concept) reported in Gaspard et al. (2018) shows, this is not necessarily the case. To account for this possibility, we therefore decided to speak of subject- and not domain-specificity in the present study.

higher academic achievement should lead to higher levels of competence beliefs. Second, students evaluate their achievement in a given domain by contrasting it with achievement they have attained in the other domain (i.e., internal reference). Internal (dimensional) comparisons are a “double-edged sword” (Möller & Marsh, 2013, p. 546) because they lead to higher self-concept within the domain that is associated with relatively higher performance while leading to lower self-concept in the domain that is associated with relatively lower academic performance. Hence, the classic I/ E model explains the paradoxical near-zero correlations between mathematics and language self-concepts through the interaction of the two counteracting processes of external and internal comparisons (Möller et al., 2009, 2020).

The GI/ E model applies the I/ E model logic to other motivational beliefs than just self-concept and to other school areas than just mathematics and languages (Möller et al., 2016). Model extensions to other school subjects were mainly guided by dimensional-comparison theory (Möller & Marsh, 2013). The theory implies that school subjects can be arranged along a continuum based on their perceived level of similarity. Mathematics and language form the opposing ends of that continuum. All other school subjects can then be positioned within this verbal-to-mathematics continuum based on their perceived similarity/ difference with both ends (Marsh et al., 1988). To the extent that two subjects are different (i.e., “far” comparisons), students are assumed to explain performance between both domains through different sets of abilities. To the extent that two subjects are seen to be similar (i.e., “near” comparisons), students are assumed to explain performance in any of the two subjects through the same set of abilities. Within far comparisons, the effects of prior achievement in one subject on motivational beliefs in the other subject are expected to be negative (i.e., contrast effects). Within near comparisons, the cross-domain effects are expected to be positive (i.e., assimilation effects; Marsh, Lüdtke, et al., 2015; Möller & Marsh, 2013). For example, Jansen et al. (2015) found assimilation effects between Chemistry and mathematics but contrast effects between Chemistry and languages. According to dimensional-comparison theory (Möller & Marsh, 2013), one would expect assimilation effects between subjects within the language domain. However, only some empirical investigations found assimilation effects (e.g., Marsh et al., 2014; Möller et al., 2006) while a large number of studies observed contrast effects (e.g., Arens et al., 2020; Marsh et al., 2001, 2014; Marsh, Lüdtke, et al., 2015; Niepel et al., 2014; Stocker et al., 2021; van der Westhuizen et al., 2022) or no effects at all (e.g., Gaspard et al., 2018; Xu et al., 2013). One explanation of these contradicting results could be rooted in the salience of languages compared. The more the two languages compared are integral parts of students' lives, as may be common in multilingual contexts, the more students are aware of their subtle nuances and thus are better positioned to recognize their differences. This in turn may then lead to contrast effects between both languages (van der Westhuizen et al., 2022).

Extensions of the classic I/ E model to other motivational beliefs concerned utility value and cost beliefs (Gaspard et al., 2018; Guo et al., 2017). Comparing different languages and quantitative subjects, Gaspard et al. (2018) found that higher school grades predicted cost perceptions favourably within domains. Between quantitative and language subjects, cost beliefs were subject to contrast effects. Between quantitative subjects, cost beliefs were subject to assimilation effects. As for utility value, they only detected within-domain effects, which were significantly weaker especially in comparison to those observed for expectancy and cost beliefs (also see Guo et al., 2017). Compared to other value facets (Götz et al., 2008; Nagy et al., 2006; van der Westhuizen et al., 2022), however, utility perceptions have been subjected to GI/E model research to a much lesser extent. More research is therefore needed to examine its relatedness to prior achievement not only in the Western world but also in Sub-Saharan Africa.

Testing the GI/ E Model within Sub-Saharan Africa

A closer look at the geographical distribution of GI/ E model studies reveals that Sub-Saharan Africa has hardly been covered (Möller et al., 2009, 2020). In fact, Chiu's (2008) study was, to the best of our knowledge, the only one that tested the I/ E model logic within the Sub-Saharan context. Using a cross-country TIMSS dataset, Chiu examined assimilation and contrasting effects on academic self-concept between mathematics and science, two related domains. She found assimilation effects in only four out of 28 contexts, including Ghana. In all other countries, including Botswana and South Africa, she found contrast effects. In her reflection of the variation within effects across geographies, she concludes that country context could constitute a moderator of the cross-domain relations between achievement and self-concept. This is in line with findings reported by Marsh and Hau (2004). Within their cross-cultural study, they found support for the classical I/ E model across 26 countries but also detected that strength of support was somewhat weaker within non-western countries (i.e., Brazil, Mexico, and South Korea). They therefore called for more empirical tests of the I/ E model logic not only within Western but also within Non-Western world, including Africa. The present study responded to this call by testing the GI/ E model's applicability within Rwanda, Sub-Saharan Africa.

A limitation of Chiu's (2008) study was that the TIMSS self-concept scales exhibited low internal consistencies especially within the three African countries. The Cronbach's alphas of the self-concept scales ranged from .41 to .56 only. In fact, it has repeatedly been highlighted that Western-borne research instruments often exhibit lower reliabilities in non-Western contexts (e.g., Bofah & Hannula, 2015; Marsh et al., 2013). It also has been argued that a reliance on Western instruments within non-Western study contexts could induce "cultural bias" (Zusho & Clayton, 2011, p. 245) into research especially if the instruments' theoretical underpinnings do not adequately reflect the socio-cultural conditions of the target context. Measurement instruments to be employed in novel research contexts have therefore been increasingly developed through bottom-up approaches (e.g.,; F. M. Cheung et al., 2011; Clayton & Zusho, 2016; R. B. King & McInerney, 2019). These approaches imply that researchers first familiarize themselves with the local meaning target population associate with a motivational construct of interest especially through the use of qualitative methods. Based on these qualitative insights, quantitative operationalizations of the construct of interest can then be developed that are attuned to the cultural-linguistic context of the target population (Zusho & Clayton, 2011). Bulla et al. (2022) provide important insights into the phenomenological structure of motivation within the Rwandan learning contexts. Using open-ended questionnaires, they collected qualitative descriptions of motivation and confidence from secondary students enrolled at two basic-education schools. Through a qualitative content analysis of the data, they found that Eccles' expectancy-value model summarized those cognitive-affective factors that Rwandan students implied within their descriptions of motivation well. As explained in section 5.2., we used both the raw data and conceptual insights of the study by Bulla et al. (2022) to derive the set of motivational self-report scales we eventually administered to the sample of the present study.

The Rwandan School Context

Rwanda is a low-income country in East Africa. Its national language Kinyarwanda is spoken by literally all Rwandans (NISR, 2016). In 2008, nine years of free and compulsory basic education (9YBE) was introduced. Students are expected to attend six years of primary education (P1 - P6) from the age of seven before moving on to three years of lower secondary schooling (S1 – S3) ideally from the age of 13. Transition into lower secondary education is conditional upon having passed national examinations at the end of primary education (Honeyman, 2015b). Their performance within these national examinations determines which schools they can attend during their secondary years. Traditionally, boarding schools have been the sole providers of secondary education in Rwanda. They continue to be regarded as “schools of excellence” (Williams, 2019, p. 648) that often reserve the right to choose those students in particular that performed well in national examinations. Thus, Rwanda’s educational system is characterized by an implicit form of school-level ability grouping (Trautwein, Lüdtke, Marsh, et al., 2006) according to which high-performers within national examinations cluster especially within more traditional school types. However, boarding schools also require tuition fees, which is not affordable especially for low-income families in Rwanda (Williams, 2019). To absorb the increasing numbers of secondary students due to 9YBE, the Rwandan government introduced basic-education schools. Even though not entirely for free (Williams et al., 2015), they are generally affordable also for low-income families. Admission to basic-education schools is also not tied to performance in national examinations. This may be one reason why schools of basic education are often perceived to be “schools for those who failed” (Williams, 2019, p. 646), an assessment not entirely fair. At times, it is just lack of resources that forces families to send their children to basic-education schools. Hence, ability grouping through basic and boarding schools may be less than perfect. Despite these inconsistencies, however, reputational differences between both school types may affect especially expectancy beliefs of basic-education students because “they compare themselves with boarding school students and feel that they are failures” (Williams, 2019, p. 651).

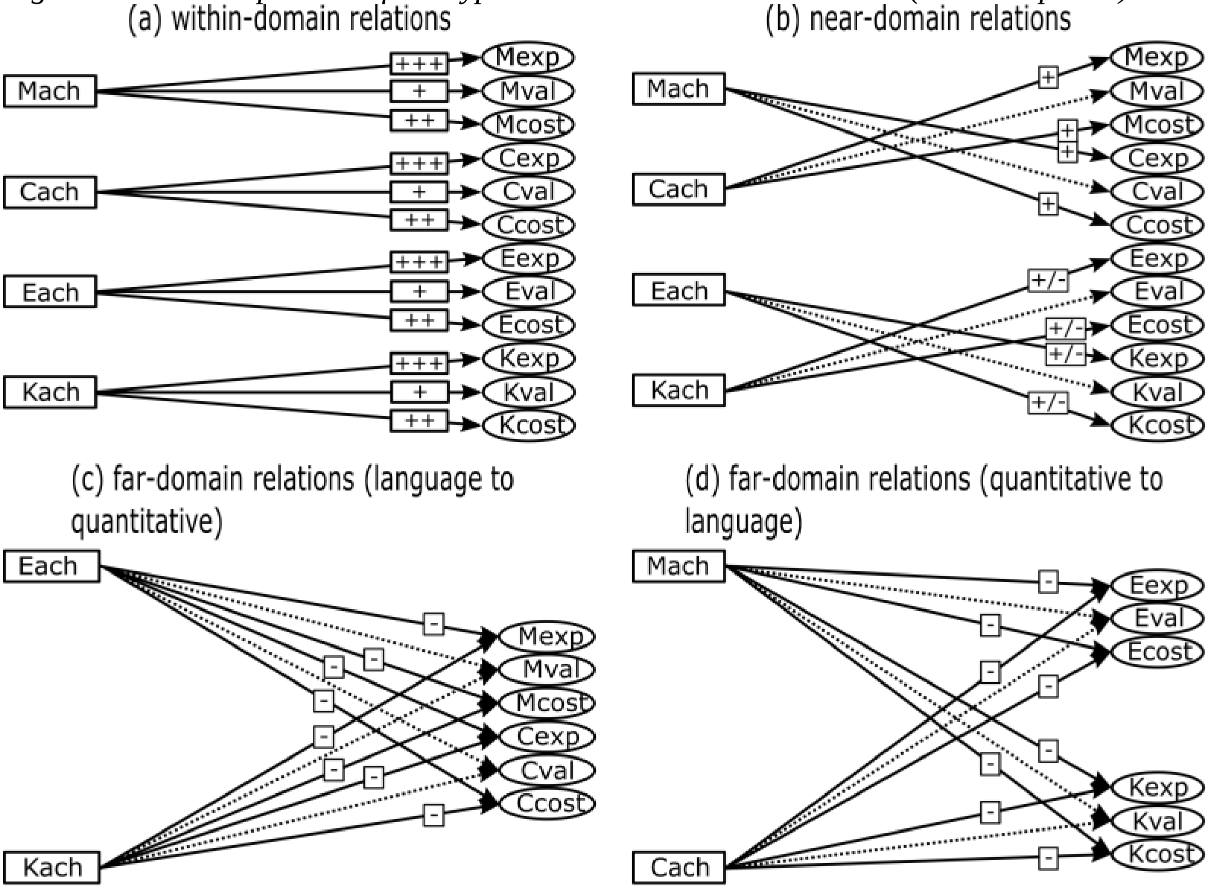
The introduction of free basic education is generally considered a success and has made Rwanda one of the top performing low-income countries in terms of getting children into schools (UNESCO, 2014). However, Rwanda has continued to struggle with learning attainment especially amongst basic-education students (Burdett & James, 2018). Poor learning outcomes have also been explained through a shift in Rwanda's medium of instruction from French to English in 2009 (Abbott et al., 2015; Honeyman, 2015a) that applied to both basic-education and boarding schools. Ever since, all school subjects but Kinyarwanda are taught in English from upper primary onwards. Hence, especially the lives of secondary students should be considered bilingual. They use Kinyarwanda especially in non-academic settings and English in academic settings. According to van der Westhuizen et al. (2022), this bilingualism should lead to contrast effects of prior achievement on motivational beliefs across English and Kinyarwanda. This may be especially true for boarding-school students. Teachers in boarding schools are considered to be more proficient in English than their colleagues at basic-education schools. Given the abrupt shift in language school policy back in 2009, especially basic-educations schools were ill-prepared for the new challenge of teaching lessons in English (Pearson, 2014; Samuelson & Freedman, 2010) something that has not changed till date (Williams, 2019). For example, more recent language assessments have found that around 40 per cent of teachers at basic-education schools still have to be considered as language beginners (Musker et al., 2014; Simpson, 2014). As a consequence, secondary-school lessons are often administered using a hybrid of Kinyarwanda and English (Williams, 2019), making the latter language potentially less salient especially in the lives of basic-education students. We therefore assumed contrast effects between English and Kinyarwanda to materialize at boarding schools but not at basic-education schools.

The Present Study

In this study, we tested the applicability of the GI/ E model to Rwandan lower-secondary students at boarding and basic-education schools. We investigated if prior academic achievement predicts students' expectancy, value, and cost beliefs across Chemistry, mathematics (both quantitative domains), English, and Kinyarwanda (both language domains). These investigations were guided by altogether five sets of hypotheses all derived from the literature reviewed above (see figure 13 for a graphical representation of the first four sets of hypotheses).

1) Due to the ability grouping within in the Rwandan education system, boarding schools are considered as 'schools of excellence' (Williams, 2019, p. 648) whereas basic-education schools are seen to be "for those who failed" (Williams, 2019, p. 646). We therefore expected that (1a) boarding-school students would exhibit higher prior school grades especially in Chemistry, Kinyarwanda, English, and mathematics than basic-education students. Due to the reputational differences between basic-education and boarding schools, we also expected that (1b) boarding-school students would hold more favourable levels of expectancy and cost beliefs across the four school subjects compared to basic-education students. However, we did not expect (1c) any differences between boarding and basic-education schools in terms of utility value.
2) We expected (2a) relations between prior achievement and motivational beliefs (expectancy, value, and costs) to be significant and positive within domains (Chemistry, mathematics, English, and Kinyarwanda). We also expected (2b) relations between achievement and expectancy beliefs to be stronger ('+++') than the ones between achievement and utility perceptions ('+') and costs beliefs ('++'). In this respect, it was expected that (2c) relations between achievement and utility value would be the weakest (e.g., Gaspard et al., 2018; also see figure 13a).

Figure 13: The components of the hypothesized structural GI/E model (manuscript two)



Note. to be tested across Chemistry, English, Kinyarwanda, and mathematics. For legibility purposes, the overall structural model is divided into four sub-components: (a) relations between achievement and motivational beliefs within matching domains; (b) relations between achievement in one subject and motivational beliefs within a different subject of the same domain (i.e., quantitative/ language). Within the language domain, the '+' refers to assimilation effects assumed to be associated with basic-education schools. The '-' refers to contrast effects assumed to be associated with boarding schools; (c) relations between achievement in the language domain (i.e., English, Kinyarwanda) and motivational beliefs within the quantitative domain (i.e., Chemistry, mathematics). (d) paths from achievement in quantitative domain (i.e., Chemistry, mathematics) to motivational beliefs within the language domain (i.e., English, Kinyarwanda). Cach = Chemistry achievement (i.e., school grades); Each = English achievement; Kach = Kinyarwanda achievement; Mach = mathematics achievement; exp = expectancy beliefs (e.g. Cexp = Chemistry expectancy beliefs); val = utility beliefs (e.g., Mval = mathematics utility beliefs); cost (e.g., Ecost = English cost beliefs). Thick lines represent hypothesized significant effects. '+++' represents hypothesized positive and substantial effects. '++' represents hypothesized positive and moderate effects. '+' represents hypothesized positive but weak to moderate effects. '-' represents hypothesized negative and weak to moderate effects. Dashed lines represent hypothesized insignificant effects.

3) We also considered to be any of the two subjects within the quantitative domain to be rather different ('far') from any of the two subjects within the language domain (and

vice versa). We therefore hypothesized (3a) that achievement in any of the two subjects of one domain (i.e., quantitative or language domains) would predict expectancy and cost beliefs within subjects of the other domain weakly but negatively (i.e., contrast effects). However, we again expected (3b) no significant effects on utility value (figure 13c and 13d).

4) We considered mathematics and Chemistry to be two similar (‘near’) school subjects within the quantitative domain (Jansen, Schroeders, et al., 2015). We therefore expected (4a) that within the quantitative domain achievement in one subject predicts expectancy and cost beliefs in the other subject weakly but positively (‘+’). In line with results reported by Gaspard et al. (2018), we did not assume (4b) achievement in one subject to predict utility value in the other subject (see figure 13b).

5) We expected assumed boarding schools to be more multilingual than basic-education schools. Due to lower multilingualism (van der Westhuizen et al., 2022), we expected (5a) weak but significant assimilation effects on expectancy and cost beliefs between English and Kinyarwanda in case of basic-education schools (which is represented through the plus sign in figure 13b). As for boarding schools, we expected (5b) weak to moderate contrasts effects of prior achievement in one language on expectancy and costs beliefs in the other language (which is represented through the minus sign in figure 13b). We did not expect (5c) any effects of achievement on utility value in the other domain.

To examine if structural relations postulated by the GI/ E model are moderated by person-level factors, we also controlled for age and gender in our analysis.

Method

Procedures

The present study was part of a wider project on achievement motivation in Rwanda that was carried out in the first two months (January and February) of the 2020 national school year around four weeks before the Covid outbreak in Rwanda. The project targeted lower secondary classes (S1 – S3). To recruit study participants, all study schools were visited during the first two weeks of the school year, which is the time traditionally given to students to register for the new academic year. On each visit, the purpose of the study was introduced, the data-collection procedure outlined, and rights of participants explained to students. To partake in the study, students signed a child assent form. Parental consent was not sought since study schools acted *in loco parentis* (British Psychological Society, 1992). As a token of appreciation, all students that submitted consent forms participated in a tombola carried out after field work. In each class, three students won a pocket calculator worth around 75p. Survey administration was handled by local research assistants all fluent in Kinyarwanda. Survey administration took around 21 minutes per class on average. To obtain measures of achievement, school grades from the previous school year (2019) were obtained from the study schools. Materials and analysis code for this study are available by emailing the corresponding author. This study was not preregistered. The study has ethics approval from University of Oxford and University of Rwanda.

Table 22: *School characteristics (manuscript two)*

School	N	Gender (in %)	Age (SE)	Che (SE)	Eng (SE)	Kin (SE)	Math (SE)
Day school (urban district)	215	50	15.39 (1.58)	40.06 (10.91)	54.2 (8.31)	37.71 (12.48)	36.64 (13.14)
Day school (rural district)	194	58	15.87 (1.56)	48.27 (8.55)	53.59 (19.05)	68.61 (8.85)	37.53 (13.15)
Day school (average)	409	54	15.62 (1.58)	43.95 (10.67)	53.91 (14.42)	52.37 (18.9)	37.06 (13.13)
Boarding school (urban district)	253	47	14.2 (1.19)	58.14 (15.4)	65.97 (9.62)	59.82 (11.56)	52.84 (15.14)
Boarding school (rural district)	107	51	14.53 (1.26)	44.97 (12.4)	54.61 (7.1)	67.66 (7.78)	51.85 (13.33)
Boarding school (average)	360	48	14.3 (1.22)	54.23 (15.75)	62.6 (10.34)	62.15 (11.16)	52.55 (14.62)
All	769	51	15.01 (1.57)	48.76 (14.24)	57.98 (13.39)	56.95 (16.49)	44.31 (15.85)

Note. ‘Basic’ refers to the school type of basic-education schools. ‘N’ stands for sample size. ‘Che’ stands for average school grades in Chemistry. ‘Eng’ stands for average school grades in English (unstandardised). ‘Kin’ stands for average school grades in Kinyarwanda (unstandardised). ‘Mat’ stands for average school grades in mathematics (unstandardised). ‘SE’ refers to standard errors. Schools A1 and A2 are from district A (urban) whereas schools B1 and B2 are from district B (rural).

Participants

Data were collected at four public schools (34 classrooms) located within two districts (one rural, one urban) of Rwanda (NISR, 2016). For each district, one boarding as well as one basic-education school was selected from the districts’ school lists. Across the 34 classrooms, 1610 students submitted valid consent forms. Out of those, 1378 students (49 per cent females) completed the survey on motivational beliefs (around 85.6 per cent). As table 22 shows, school grades for Chemistry, English, Kinyarwanda and mathematics from the previous school year were available for 769 of those 1378 students (around 54.8 per cent). They were mainly from S2 (48.9%) and S3 (48.4%) and only to small extent from S1 grades

(2.7%). In 2019, most S1 had still been in primary schools, which tended to be different from those schools they started their secondary schooling at the time of data collection. It was therefore impossible to obtain their 2019 school grades in most cases. The average number of students per class for which achievement data were available was 24.8 ($SD = 15.8$) and ranged from 1 to 34 students. Table 22 provides additional information on sample characteristics. Students from basic-education schools were more likely to be female and older. Compared to all other schools, boarding school A1 performed significantly better in Chemistry, and English. Comparing school-grade averages, boarding schools generally outperform basic-education schools.

Measures

Motivational Beliefs

Students reported their motivational beliefs across the four subject areas by using one of four versions of the study questionnaire, all in Kinyarwanda. The versions only differed in terms of the order of the four modules in which students were asked to report their beliefs (e.g., English, Chemistry, Kinyarwanda, mathematics). Each version started by asking personal information (e.g., name, age, gender) from students. Scales to measure student motivation within the Rwandan context had not existed prior to study implementation to the best of our knowledge. We therefore carried out a robust scale-development process to develop cultural-linguistically sensitive measures needed to gauge motivational beliefs amongst Rwanda students. The starting point of this process was the qualitative study reported by Bulla et al. (2022). It provided insights into phenomenological structure of Rwandan students' motivational experiences within learning contexts. The themes within student descriptions were adequately summarized through an expectancy-value-cost coding frame. The emerging themes were then translated into a provisional set of motivational scales. Students' way of wording to express motivational experiences within the raw data collected by Bulla et al. (2022), were used to develop the corresponding self-report items. These provisional scales were then evaluated through cognitive interviews and two rounds of exploratory factor analyses (EFAs). The scale-development process, outlined in section 5.2, eventually resulted in three motivational scales: one on expectancy beliefs, one on utility perceptions, and one on cost beliefs. Due to the data-driven approach towards scale development, however, the scales slightly varied across the quantitative (Chemistry and mathematics) and language (English and Kinyarwanda) domain, something not entirely uncommon within GI/ E model research (e.g., Arens & Preckel, 2018; Guo et al., 2018). Within the quantitative domain, expectancy beliefs were measured through seven items, value through eight (Chemistry) and seven (mathematics) items[20], respectively. Cost beliefs were measured through six items. Within the language domain, expectancy and utility beliefs were gauged through six items each, and cost beliefs through eight items (see appendix S5 in the supplemental material for motivational scales). Given learnings from cognitive interviews (see section S1 of the supplemental material), all items used a three-answer response format: 1) often; 2) sometimes; or 3) never. Costs items were reverse coded and thus higher value reflect more favourable cost beliefs. Table 23 presents average polychoric correlations across items within each scale and domain as a measure of reliabilities.[21] Motivational scales attained higher reliabilities amongst boarding-school students compared to basic-education

[20] Initially, maths value was also modelled through eight items. In some models, however, the eight-item scale resulted on non-positive definitive covariances matrices. We therefore removed this item not only in those models but across all models. As a result, maths daily utility was modelled only through seven items.

[21] Within the literature, no consensus has yet been reached on how to compute reliabilities of scales involving categorical items with more than two items. It is for this reason that we relied on polychoric correlations to obtain at least a crude measure of reliability rather than any of the other, commonly used but necessarily more reliable measures.

students. To determine whether variability within scales between subjects influenced results, we rerun key models with scales trimmed down to those items that were used in all four subjects (see section 3.6).

Table 23: *Average within-scale polychoric correlations (manuscript two)*

Motivational scale	Item statistics				Average
	Che	Eng	Kin	Mat	
Expectancy	7/ 0.48/ 0.58	6/ 0.71/ 0.74	6/ 0.69/ 0.80	7/ 0.69/ 0.74	0.64/ 0.72
Value	8/ 0.64/ 0.71	6/ 0.59/ 0.58	6/ 0.58/ 0.68	8/ 0.52/ 0.51	0.58/ 0.62
Cost	6/ 0.50/ 0.62	8/ 0.51/ 0.54	8/ 0.58/ 0.55	6/ 0.32/ 0.42	0.48/ 0.53
Average	0.54/ 0.64	0.60/ 0.62	0.62/ 0.68	0.51/ 0.56	0.57/ 0.62

Note. ‘Che’ stands for Chemistry. ‘Eng’ stands for English. ‘Kin’ stands for Kinyarwanda. ‘Mat’ stands for mathematics. Within each cell, the integer refers to the number of items within a scale. The first fraction depicts the average polychoric correlations between items within a scale for the case of basic-education schools whereas the second fraction depicts the corresponding average for the case of boarding schools.

Achievement

Academic achievement was operationalised through final school grades from the previous school year (e.g., Gaspard et al., 2018). Students are used to school grades as measures of academic progress. School grades may therefore be more predictive of subsequent motivational beliefs than, for example, one-time external learning assessments (Marsh, 1987; Möller et al., 2009). In Rwanda, end-of-year school marks are generally presented as per centages with 100 per cent indicating perfect achievement. End-of-year marks represent averages students attain across the three school terms that make up the Rwandan school year. First- and second-term grades are generally determined through school exams and tests organized locally. School exams within the third term are prepared by the national level and uniformly administered across schools in the country (Nzabalirwa, 2014). It was also for this reason that we considered end-of-year school marks as somewhat comparable across schools. We standardised school grades since it allowed estimating average effect sizes of achievement within and across domains for each of the different motivational beliefs in a more meaningful way (see bottom of table 26). Given this unequal sample size across school grades, we also decided to run a secondary model without S1 students.

Statistical Analysis

We used Mplus 8.7 (L. K. Muthén & Muthén, 2017) to perform the statistical analysis. Achievement data was generally normally distributed with an average skewness of -0.08 as well as average kurtosis of 2.87 (see S10 within the supplemental materials for more information). Amongst motivational items, the average skewness was -0.87 while the average kurtosis was 3.31. None of the items exhibited skewness exceeding 2 and almost all items but one on daily utility exhibited kurtosis beyond 6. Nonetheless, due to the ordinal nature of motivational data (i.e., three response options), all latent models were estimated using robust weighted least squares (WLSMV). At the classroom level, motivational data exhibited low levels of dependency with an average intra-class correlation (ICC) of 0.04 and a maximum of ICC of 0.10 (see S9 within the supplemental materials for more information) whereas achievement data was rather clustered with an average ICC of 0.58 (see S11 within the supplemental materials for more information). Patterns of low ICCs within motivational data with high ICCs in achievement data is not uncommon (e.g., Trautwein, Lüdtke, Marsh, et al., 2006). To account for nesting of students within classrooms, design-based correction of standard errors were used through the “type=complex” option within Mplus (McNeish et al., 2017). Missing data was handled through the full information maximum likelihood procedure (i.e., FIML; Enders, 2010; J. W. Graham, 2009). Amongst motivational items, missingness ranged from 1 to 6% of the data (average 2% per person or per item). Given that we only focused on students whose school grades were fully available, there was no missingness within school grades.

To analyse the data, we first run separate confirmatory factor analyses (CFA; Brown, 2015) for both the basic-education and the boarding-school subsample (model 1 and 2). Through imposing a latent factor structure consisting of expectancy, utility-value, and cost beliefs in each of the four domains, we further evaluated the adequacy of the measurement scales that were first tested through the two rounds of EFAs (see supplemental material). Through model 3 and 4, we tested the applicability of the GI/ E model to students' expectancy, daily utility, and cost beliefs separately within basic-education and boarding schools. Both models provided first evidence on the adequacy of the GI/ E model logic to both school types. To assess differences within the applicability of the GI/ E model to both school types, we then performed multigroup analysis of the combined sample within subsequent models. Within model 5 and 6, we tested for invariance of measurement models across school type. Since all scales were categorical in nature, measurement invariance was tested largely following the approach proposed by Wu and Estabrook (2016) and further elaborated by Svetina et al. (2020). According to this approach, one would first estimate model in which all thresholds and loadings were estimated freely across groups (configural invariance). In subsequent models, first thresholds (threshold invariance) and then both thresholds and loadings (threshold and loading invariance) are then constrained to be equal across groups. In our case, configural-invariance and threshold-invariance models were equivalent because we employed motivational scales with three answer options only (H. Wu & Estabrook, 2016). Items with only three answer options only have two latent thresholds. Within the configural-invariance model, these thresholds are estimated freely across the two groups. Within the threshold-invariance model, thresholds are constrained to be equal across groups. Instead, intercepts and residual variance associated with the items are estimated freely within the comparison model. Thus, both models do not differ in terms of degrees of freedom. In those cases, threshold invariance can therefore only be assumed but not tested. The corresponding threshold-invariance model (i.e., model 5) was thus treated as the baseline model whereas model 6 (i.e., threshold- and loading-invariance model) was the final model within the measurement-invariance testing process. Upon establishment of measurement invariance across school3types, we then respecified model 6 to be able to extract school type-specific factor correlations (see table 24). To do so, we constrained factor variances equal to 1 not only in the reference group (i.e., basic-education schools) but also within the comparison group (i.e., boarding schools). As highlighted in table 24, fit hardly deteriorates through this implicit assumption of equal variance across groups. Through models 7 to 10, we then tested the GI/ E model in which the different motivational factors were regressed on achievement within each of the four domains using multigroup structural equation modelling (SEM; e.g., Bollen, 1989). The final model 9 was used to test the most hypotheses guiding this study. It provided information on paths leading from achievement to motivational beliefs within matching domains (figure 13a), from achievement to motivational beliefs within non-matching but near domains (figure 13b) as well as to motivational beliefs within non-matching, far domains (figure 13c and 13d) for the case of both basic-education and boarding schools. To build up the final model, we again proceeded in a way similar to the one outlined for CFAs. In Model 7, we combined both sub-samples and estimated the structural model with thresholds constrained to be equal across school type. We then compared this model with model 8 in which both thresholds and factor loadings were constrained to be equal across groups.[22] Since invariance held, the first factor loading of each factor was then taken and included into the final model 9 to test the GI/E model (common metric standardization).

[22] Models 7 and 8 are respecifications of models 5 and 6 because latent correlations involving motivational factors and school grades are turned into latent regressions of motivational factors on school grades. Hence, models 5 and 7 as well as 6 and 8 each depict the same statistical fit as highlighted in table 24.

To estimate average effect sizes associated with within-domain and cross-domain paths for each of the different motivational beliefs (see bottom of table 26 on page 204), we used Mplus' model-constraint option. Model 10 was estimated to determine whether or not path coefficients differed across school type. We also run additional models which controlled for gender and age (see annex 8), which excluded S1 students (see annex 8), and which included only those self-report items that appeared in all four domains (see annex 8). Because the chi-square (χ^2) statistic is sensitive to sample size (G. W. Cheung & Rensvold, 2002; Dimitrov, 2010), model fit was evaluated using the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR; e.g., Marsh et al., 2004). As for CFI and TLI, values above 0.90 and 0.95 represent adequate respectively good model fit (Hu & Bentler, 1999). As for RMSEA, values below 0.08 and 0.05 represents a reasonable respectively good model fit (Browne & Cudeck, 1993). SRMR below 0.08 is seen to indicate good model fit (Hu & Bentler, 1999). Non-invariance was concluded when changes in CFI (ΔCFI) and in RMSEA (ΔRMSEA) between model 5 and 6 as well as between model 7 and 8 were equal or greater than -.010 and less than or equal to .015, respectively (Chen, 2007). We also specified correlated uniqueness between parallel worded cost items across domains (Marsh et al., 2013; Marsh & Hau, 1996) and that way accounted for reverse coding of these items.[23]

[23] In line with research practice (Marsh et al., 2013; Marsh & Hau, 1996), we initially specified correlated uniqueness between parallel worded items for all motivational scales. However, this specification partly resulted in non-positive definitive covariance matrices during model estimations. We therefore decided against specifying correlated uniqueness between parallel worded items across domains within expectancy and value scales.

Table 24: *Goodness of model fit (manuscript two)*

Model	descriptions	N	df	χ^2	RMSEA	CFI	TLI	SRMR	Comparison	ΔCFI	ΔRMSEA/ ΔSRMR
M1	CFA with basic-education subsample	409	3336	3555.0	.013	.959	.956	.072			
M2	CFA with boarding-school subsample	360	3336	3523.8	.013	.972	.970	.072			
M3	SEM to test GI/E model with basic-education subsample	409	3336	3566.4	.013	.956	.953	.071			
M4	SEM to test GI/E model with boarding-school subsample	360	3336	3536.8	.013	.961	.958	.073			
M5	Multigroup CFA with thresholds invariance	769	6672	7084.6	.013	.966	.964	.072			
M6	Multigroup CFA with thresholds and loading invariance	769	6740	7155.1	.013	.966	.964	.072	5 vs. 6	-.000	.000/.000
M7	Multigroup SEM of GI/E model with thresholds invariance	769	6672	7084.6	.013	.966	.964	.072			
M8	Multigroup SEM of GI/E model with thresholds and loading invariance	769	6740	7158.2	.013	.966	.964	.072	7 vs. 8	-.000	.000/.000
M9	Multigroup SEM of GI/E model with common-metric standardization	769	6740	7177.3	.013	.958	.955	.072			
M10	Model 9 with path-coefficient invariance	769	6788	7268.4	.014	.954	.951	.079	9 vs 10	-.004	.001/.007

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardised root mean squared residual. All $\chi 2$ values are significant ($p < .0001$).

Table 25: *Standardised factor correlations (manuscript two)*

	Chemistry				English				Kinyarwanda				Mathematics			
	1 AC	2 EX	3 DV	4 CO	5 AC	6 EX	7 DV	8 CO	9 AC	10 EX	11 DV	12 CO	13 AC	14 EX	15 DV	16 CO
1	-	.47***	.24***	.44***	.62***	-.10*	.08	.07	.45***	-.10	.01	-.05	.60***	.21*	.22***	.24***
2	.23***	-	.61***	.71***	.24**	.18***	.01	.12*	.02	.03	.13	.06	.21*	.44***	.32***	.37***
3	.15***	.65***	-	.48***	.11*	.17**	.25**	.18***	.13*	.09	.28**	.10	.10	.24***	.56***	.25***
4	.09**	.46***	.28***	-	.19*	.10	-.02	.36***	.10	.09	.23**	.30***	.19*	.34***	.26***	.57***
5	.36***	.02	.04	.04	-	.30***	.16*	.30***	.21**	-.18***	-.03	-.06	.40***	-.01	.19***	.10
6	.02	.39***	.24***	.08	.03	-	.53***	.61***	-.16***	.23***	.17**	.15**	-.29***	.02	.23***	.11
7	.15**	.19**	.55***	.02	.08	.47***	-	.32***	.10**	.19***	.45***	.14***	.05	.08	.36***	.09*
8	.00	-.02	.08	.28***	.16**	.44***	.32***	-	-.11	.11*	.11	.40***	-.09	-.04	.15**	.29***
9	.45***	.14*	.14*	.00	.08	.07	.13***	.08*	-	.36***	.21***	.19**	.39***	.06	.13	.04
10	-.02	.22***	.14***	.03	.01	.32***	.18**	.12**	.19***	-	.47***	.70***	-.08	.12	.15**	.07
11	.07	.19***	.40***	.05	.03	.12*	.37***	.04	.11	.47***	-	.33***	.01	.18**	.36***	.21**
12	-.07	-.12**	-.08	.27***	.09	.01	.03	.54***	.06	.35***	.10	-	-.04	.07	.14***	.35***
13	.39***	.19***	.13*	.16***	.03	-.01	.02	-.04	.09	-.04	.05	.01	-	.49***	.21***	.37***
14	.18***	.52***	.35***	.26***	.00	.07*	.06	-.06	-.01	.13**	.17*	-.05	.50***	-	.42***	.69***
15	.17**	.33***	.57***	.15***	.00	.06*	.49***	.02	.08**	.10*	.37***	.04	.21***	.53***	-	.35***
16	.08*	.15***	.12*	.59***	.07	-.08	-.09	.11**	-.08	.04	-.02	.30***	.31***	.50***	.22***	-
Δ Mean	10.3**	.52***	.10	.17	8.7**	.96***	.19	.39**	9.8*	-.31*	-.26**	-.09	15.5***	.54***	.24*	.27**

Note. Cost-beliefs were reverse coded, higher values indicating more favourable cost beliefs. Correlations were standardised across groups by fixing factor variances to 1. Basic-education schools below the diagonal/ boarding schools below the diagonal. AC = achievement; EX = expectancy; DV = daily utility; CO = costs. 'Δ Mean' refers to differences in factor means between boarding and basic-education schools. Fit of CFA to extract latent mean differences: χ^2 (df = 6752) = 7173.7, RMSEA = .013, CFI = .965, TLI = .963, SRMR = .073; ***p < .001, **p < .01, *p < .05.

Results

Invariance Tests and factor correlations

As table 24 highlights, the hypothesized factor structure imposed on the motivational scales across the four school subjects fitted the data adequately amongst basic-education (model 1) and boarding schools (model 2). We took this as further evidence in support of the hypothesized factor structure associated with scales that emerged from scale-development results (see supplemental material). Model 3 and 4 tested the applicability of the GI/ E model to both school types. Model fit was again acceptable in both cases. Path coefficients with the structural GI/E model applied to both subsets are presented in annex 8. Overall, results resemble the ones obtained during the main analysis of model 9. We therefore refrain from a discussion of the subset-specific findings here. Through model 5 and 6, measurement invariance across school type was tested. To do so, both subsamples were combined within model estimation. Model 5 fit the data well, demonstrating that the overall factor structure and corresponding threshold invariance describes the data well. Within model 6, factor loadings were constrained to be equal across groups. Fit indices did not deteriorate at all, which suggests that the more restricted model with factor invariance fits the data well. Hence, both threshold and factor-loading invariance was concluded across all setups (Chen, 2007). Through constraining factor variances within both subsamples within model 6, we were then able compare factor correlations across both groups - that is, to test whether the basic-education schools and boarding schools differed in terms of factor correlation patterns. Table 25 presents the factor correlations. Amongst boarding schools, achievement was more strongly correlated across domains ($\bar{\rho}_{BOA}$ = .56; SE = .06) than amongst basic-education schools ($\bar{\rho}_{BAA}$ = .40; SE = .17). For example, achievement was substantially correlated between mathematics and Chemistry as well as within languages. By contrast, mathematics was almost uncorrelated with both English and Kinyarwanda. Amongst boarding schools, motivational factors of the same type correlated slightly weaker between domains ($\bar{\rho}_{BO}$ = .31; SE = .021) than in the case of basic-education schools ($\bar{\rho}_{BA}$ = .36; SE = .018). In comparison to utility and cost perceptions, expectancies exhibited the highest levels of domain-specificity of the three motivational variables amongst both basic-education ($\bar{\rho}_{BAE}$ = .28; SE = .017) and boarding schools ($\bar{\rho}_{BOE}$ = .17; SE = .023). However, compared to what has been reported in the literature (e.g., Gaspard et al., 2018), expectancy beliefs still exhibited higher cross-domain correlations within the Rwandan context. This in turn could be interpreted as competence beliefs being relatively domain-general amongst Rwandan students. Amongst basic-education schools, daily utility exhibited the highest correlations across domains ($\bar{\rho}_{BAU}$ = .46; SE = .036), which was higher than respective cost beliefs ($\bar{\rho}_{BAC}$ = .35; SE = .023). Amongst boarding schools, cross-domain correlations of daily-utility ($\bar{\rho}_{BOU}$ = .38; SE = .042) and cost beliefs ($\bar{\rho}_{BOU}$ = .38; SE = .025) were at par.

Table 25 also presents mean differences within achievement and motivational beliefs between basic-education and boarding schools. A comparison of unstandardised school grades suggests that boarding-school students generally perform better than basic-education students. Boarding-school students also exhibited higher expectancy beliefs in Chemistry, mathematics, and in particular English. Within English, expectancy beliefs were higher by almost an entire standard deviation. Boarding-school students also exhibited more favourable cost beliefs in English and mathematics while showing lower levels of motivational beliefs only within Kinyarwanda. The latter may be an indication of the dominance of Kinyarwanda within learning contexts of boarding schools. Basic-education and boarding school students did not differ in terms of their value beliefs.

Table 26: *Estimated coefficients of the structural models (manuscript two)*

Predictor	Expectancies		Value		Costs		Predictions
	Basic	Boarding	Basic	Boarding	Basic	Boarding	
Chemistry outcomes							
Chemistry	0.191*	0.665***	0.078	0.314**	0.006	0.627***	Match (++)
English	-0.069	-0.108	0.003	-0.072	0.026	-0.159	Far (-)
Kinyarwanda	0.053	-0.224***	0.100	0.023	-0.003	-0.131**	Far (-)
Mathematics	0.120	-0.062	0.093	-0.072	0.152*	-0.074	Near (+)
English outcomes							
Chemistry	-0.012	-0.237***	0.078	-0.068	-0.139*	-0.004	Far (-)
English	0.027	0.588***	0.043	0.199***	0.216***	0.359***	Match (++)
Kinyarwanda	0.070	-0.030	0.108*	0.098	0.133***	-0.106	Far (-)
Mathematics	-0.013	-0.391***	-0.039	-0.062	0.006	-0.199*	Far (-)
Kinyarwanda outcomes							
Chemistry	-0.151*	-0.217***	0.035	-0.046	-0.211***	-0.165	Far (-)
English	0.049	-0.137*	0.008	-0.050	0.152**	-0.000	Far (-)
Kinyarwanda	0.250***	0.536***	0.103	0.249***	0.139***	0.267***	Match (++)
Mathematics	-0.002	-0.157*	0.028	-0.057	0.066	-0.049	Far (-)
Mathematics outcomes							
Chemistry	0.037	0.085	0.088	0.111	-0.045	0.155*	Near (+)
English	-0.009	-0.252***	-0.042	0.058	0.058	-0.112	Far (-)
Kinyarwanda	-0.063	-0.156***	0.020	0.021	-0.103***	-0.133**	Far (-)
Mathematics	0.501***	0.592***	0.172**	0.083	0.339***	0.400***	Match (++)
Summary (across the different set of paths)							
Matching paths							
M	0.289***	0.828***	0.102*	0.258***	0.187**	0.501***	
# sign. Paths	3	4	1	3	3	4	
Nonmatching paths							
M	0.000	-0.217***	0.036	-0.012	0.015	-0.097***	
# sign. Paths	1	8	1	0	6	4	
Nonmatching (near) paths							
M	0.075*	-0.050	0.077**	0.027	0.101***	-0.004	
# sign. Paths	0	1	1	0	2	1	
Nonmatching (far) paths							
M	-0.037	-0.300***	0.023	-0.032	-0.039*	-0.143***	
# sign. Paths	1	7	0	0	3	3	

Note. Cost-beliefs were reverse coded, higher values indicating more favourable cost beliefs. All parameters are standardised. Within each construct, the first coefficient refers to subsample of basic-education schools (Basic). The second coefficient refers to the boarding-school subsample (Boarding). ***p < .001; **p < .01; *p < .05.

In order to compare the applicability of the GI/ E model to both basic-education and boarding schools, we first established invariance within SEMs across school type (model 7 and 8). The baseline model 7 with threshold invariance again showed adequate fit. Also, constraining factor loadings across school type did not entail any deterioration in model fit 6see table 24). Hence, both threshold and factor-loading invariance was concluded across all domain-specific setups (Chen, 2007). Using common-metric standardization, we then estimated model 9 to test the applicability of the GI/ E model to both basic-education and boarding schools. Table 26 provides the resulting path coefficients, based upon which research questions one through five will be answered in the following sections.

Within-domain effects of achievement on motivational beliefs

Amongst boarding schools, 11 out of 12 effects on motivational beliefs within domains were statistically significant. Only mathematics utility was not predicted by mathematics school grades. Average effect sizes were the strongest amongst expectancy beliefs ($\bar{\beta}_{BOE}$ = .828; SE = .069). By comparison, effects on cost ($\bar{\beta}_{BOC}$ = .501; SE = .054) and utility perceptions ($\bar{\beta}_{BOU}$= .258; SE = .044) were significantly smaller. In case of basic-education schools, seven out of 12 effect sizes were statistically significant. Expectancy beliefs in all domains but English were predicted by matching school grades. Cost beliefs in all domains but Chemistry were predicted by prior achievement. Within daily utility, respective perceptions were predicted by matching school grades only within the mathematics domain. Average within-domain effect sizes were also the smallest in the case of daily utility ($\bar{\beta}_{BAU}$ = .102; SE = .034) compared to cost ($\bar{\beta}_{BAC}$ = .187; SE = .017) and especially expectancy beliefs ($\bar{\beta}_{BAE}$ = .289; SE = .032). Hence, within-domain effects of prior achievement on expectancy (β_{Diff_Ex} = .538; SE = .075), utility value (β_{Diff_UV} = .157; SE = .054), and cost beliefs (β_{Diff_Co} = .314; SE = .056) were significantly larger amongst boarding schools than basic-education schools. This was particularly true in the case of expectancy and cost beliefs.

Cross-domain effects between language and quantitative subjects

We also observed differences between basic-education and boarding schools in terms of effects between far domains (i.e., language versus quantitative subjects). Amongst boarding schools, seven out of eight of effects on expectancy beliefs were statistically significant and negative as predicted by theory. Amongst cost beliefs, it was three out of eight paths. Far-domain effects on expectancy beliefs (M= -0.301; SE = .032) were larger than the ones on cost beliefs (M= -0.143; SE = .024). By contrast, average far-domain effects on utility value were insignificant (M = -0.032; SE = .031). Amongst basic-education schools, far-domain effects were less common. They mainly appeared amongst cost beliefs. Cost beliefs on mathematics were negatively predicted by Kinyarwanda grades whereas both English and Kinyarwanda cost beliefs were negatively predicted by Chemistry grades. Within expectancy, we observed the far-domain effect of Chemistry achievement on Kinyarwanda competence-beliefs. Average effect sizes were only significant in case of cost beliefs (M = -0.039; SE = .020).

Cross-subject effects within the quantitative domain

We also examined cross-path effects between subjects considered to be similar. Within the quantitative domain, this concerned effects between Chemistry and mathematics. In case of boarding schools, we did not find any cross-domain effects of prior achievement on motivational beliefs with the sole exception of mathematics cost beliefs. The effects of Chemistry school grades were positive and statistically significant. Amongst basic-education schools, relations were in reverse. Here, we found mathematics school grades to positively predict Chemistry cost beliefs. Besides, however, we found no further cross-subject effects within the quantitative domain.

Cross-subject effects within the language domain

In addition to cross-subject effects within the quantitative domain, we also examined respective effects within the language domain. Patterns again differed between basic-education and boarding schools. In the case of the former, we found assimilation effects especially amongst cost beliefs. Between English and Kinyarwanda, we found prior achievement in one subject to positively influence cost beliefs in the other subject. As table 26 highlights, these cross-subject effects were not much weaker than the within-subject effects of prior achievement. Within expectancy beliefs, we did not find any cross-subject effects between both languages for the case of basic-education schools. However, we found Kinyarwanda school grades to positively predict English daily utility, it was one of two significant effects we found amongst daily utility reported by basic-education school students. Amongst daily utility reported by boarding-school students, we found neither any assimilation nor contrast effects. In fact, besides English school grades to negatively predicting Kinyarwanda expectancy beliefs, we found no cross-subject effects within the language domain amongst boarding schools.

The moderating effects of age and gender

In addition to those models presented in table 24, we also rerun model 9 using different parameterizations. First, we rerun the model controlling for gender and then for age (see annex 8). The respecification left the structural relations reported above unchanged by and large. Within the model controlling for age, only the coefficient of Kinyarwanda achievement on Chemistry cost beliefs amongst boarding-school students turned insignificant. Within the model controlling for gender, only the coefficient of Chemistry achievement on Kinyarwanda expectancy beliefs amongst basic-education students turned insignificant. Gender was not always a statistically significant predictor of motivational beliefs. Nonetheless, it appeared that female students generally tend to have lower expectancy beliefs, higher value perceptions, and more negative cost beliefs, findings that are in line with has been reported in the literature (e.g., Else-Quest et al., 2010a). Age was not always a significant predictor either. But similar to gender, older students tend to have lower expectancy beliefs in Kinyarwanda (basic-education school) and Chemistry (boarding schools), and more negative cost beliefs in Chemistry (boarding schools), English (basic-education school), and Kinyarwanda (basic-education and boarding schools).

Second, we rerun the model using only those self-report items employed in all four subjects. As mentioned above the composition of motivational self-report scales slightly varied between the language and quantitative domains. To check if this variability impacted structural relations, we rerun model 8 and 9 using only those 17 items across the three scales that were used in all four domains. Using $p < 0.5$ as a threshold, this respecification resulted in fewer significant effects especially in the case of expectancy beliefs (see annex 8). Relaxing the level of significance to $p < .10$, however, resulted on patterns similar to those reported in table 26. We took this as evidence that results in table 26 are robust despite the variability within scales between subjects.

Discussion

In our study, we examined the applicability of the GI/ E model to explain motivational beliefs exhibited by lower-secondary students from four public schools in Rwanda. Given its geographical focus, our study responded to Marsh's and Hau's (2004) call for additional research on the I/ E model logic also within the African context. Our investigation also went beyond Chiu's (2008) cross-country study might have been the first GI/ E model research in African contexts. First, through integrating the generalized I/ E model (Möller et al., 2016) with expectancy-value theory (e.g., Eccles et al., 1983; Eccles & Wigfield, 2002), our study investigated internal and external frame-of-reference effects of prior academic achievement not only on expectancy beliefs but also daily-utility and cost perceptions. Only few studies (e.g., Arens et al., 2017; Gaspard et al., 2018; Wolff et al., 2018) within GI/ E model research focused on cost beliefs. Second, unlike Chiu (2008) we did not only focus on science (i.e., Chemistry and mathematics) but also covered the language domain by including English and Kinyarwanda, the former Rwanda' s main medium of instruction and the latter being Rwanda's national language. By contrasting both languages, the study contributed a first African perspective to the discourse on whether or not different language subjects constitute similar or dissimilar school domains. Third, we controlled for the fact that ability grouping of students is a common practice within many African educational systems. In our study, we compared the applicability of the GI/ E model to motivation formation at both Rwandan public boarding schools as well as basic-education schools. Admission to the former is often tied to high-performance at national examinations whereas the latter are free for everyone. Fourth, whereas Chiu (2008) employed motivational scales from the Western context (i.e., TIMSS), we developed self-report instruments that were attuned to the cultural-linguistic context of Rwanda, which in turn should contribute to the ecological validity of study results. Below, we discuss study results on the applicability of the GI/ E model in light of the different hypotheses that guided above mentioned data analysis.

Do boarding and basic-education schools differ in terms of achievement and motivational levels (hypotheses 1a, 1b and 1c)?

As highlighted in table 25, boarding-school students performed significantly better in all four subjects than basic-education students, which confirmed hypothesis 1a. Boarding-school students also exhibited higher expectancy beliefs within Chemistry, English, and mathematics, which was in line with hypothesis 1b. Boarding-school students also reported more favourable English and mathematics cost beliefs as well as mathematics utility perceptions. In Kinyarwanda, however, boarding-school students exhibited both lower expectancy and value beliefs compared to basic-education students. Lower expectancy beliefs could already have been an indication of the contrast effect we hypothesized to exist between English and Kinyarwanda in case of boarding schools. After all, boarding-school students tend to have a better command over English than basic-education students as well. In terms of Chemistry, and English daily utility, boarding and basic-education students did not differ, which we considered largely to be in line with hypothesis 1c. An interesting future research question in this respect would be if students from different school types but with a similar level of prior academic achievement differ in terms of their expectancy and cost beliefs. According to the proposed the big-fish-little-pond effect (e.g., Trautwein, Lüdtke, Marsh, et al., 2006), respective students from boarding schools should express lower expectancy (and cost?) beliefs compared to basic-education school students due to the high-performance environment associated with the former school types.

Does achievement predict motivation within domains (hypotheses 2a, 2b, and 2c)?

The applicability of the GI/E model to students' motivational beliefs varied across school types. The GI/ E model seemed to explain motivational data collected at boarding schools quite well. In line with model predictions, boarding-school students' prior achievement was substantially related to expectancy, value, and cost beliefs within domains, something that has been observed in other contexts as well (e.g., Gaspard et al., 2018). The only exception we found was mathematics value that appeared not to be influenced by mathematics achievement. Nonetheless, we considered hypothesis 2a of positive effects of achievement on all within-domain motivational beliefs confirmed in the case of boarding schools. Since average effect sizes related to expectancy beliefs ($\bar{\beta}_{BOE}$= .828; SE = .069) were the strongest while average effect sizes related to utility beliefs ($\bar{\beta}_{BOU}$ = .258; SE = .044) were the smallest, we also considered hypotheses 2b and 2c confirmed in case of boarding schools. In case of basic-education schools, hypothesis 2a was largely confirmed. Almost all expectancy ($\bar{\beta}_{BAE}$= .289; SE = .032) and cost beliefs ($\bar{\beta}_{BAC}$= .187; SE = .017) were determined by prior school grades within the matching domain. Conversely, utility-value perceptions ($\bar{\beta}_{BAU}$= .102; SE = .034) were mainly unrelated to prior achievement, which in turn could be taken as support of hypothesis 2c. Average effects on expectancy beliefs were larger than average effects on cost beliefs, which was in support of hypothesis 2b in case of basic-education students.

The partly strong associations between prior achievement and expectancy (boarding schools) and cost beliefs (basic-education and boarding schools) within matching subjects implies that social comparisons are important in explaining student motivation. In fact, Rwanda's school system socializes students to compare themselves in terms of their performance. School report cards inform students not only about school subject-specific school marks but also about the relative ranking they attained based on their overall performance in relation to other students in class. Moreover, report cards are generally given out to students based on their relative ranking clearly highlighting who are the worst and best performers in class. Moreover, at the end of each school year the three top performers of each class are publicly honoured by the headmaster in front of all students. This all may instil the habit of social comparisons amongst Rwandan students. Social comparisons also appear to shape daily-utility perceptions of school subjects in case of boarding but not basic-education schools. The latter aspect coincides with the fact that respective value beliefs amongst basic-education students exhibited the lowest level of domain-specificity of all motivational beliefs across both school types. As table 25 shows, daily utility within basic-education schools showed the highest average cross-domain correlation ($\overline{\rho_{DV}}$ = .44) of all motivational beliefs across both school types ($\overline{\rho_{rest}}$ = .31). As explained above, domain-specificity is a key precondition for the applicability of the GI/E model of belief formation (Möller et al., 2016). Average effect sizes of achievement on expectancy and cost beliefs were substantially larger amongst boarding than basic-education schools. One explanation could be that reputational differences between both school types moderate effects of prior achievement. After all, basic education schools are seen to be for those “who failed” (Williams, 2019, p. 646). This may entail that basic education-school students do not develop the same motivational levels in response to prior academic success as boarding school students may do. Future research is required to determine if this hypothesis indeed holds true.

Does achievement predict motivation between subjects that each belong to different (far) domains (hypotheses 3a, and 3b)?

In case of boarding schools, far-domain effects were common especially within expectancy beliefs (seven out of eight paths significant) and somewhat common within cost beliefs (three out of eight paths significant). We therefore considered hypothesis 3a, which postulates far-domain effects of prior achievement on respective beliefs only partly supported in the case of boarding schools. Amongst basic-education schools, far-domain effects were hardly observed within expectancy beliefs (only one out of eight paths significant) but were somewhat more common within cost beliefs (three out of eight paths significant). We therefore considered hypothesis 3a unsupported in the case of basic-education schools. Overall, we therefore found the most fundamental aspect of the GI/ E model logic to be at work only in case of boarding-school students' expectancy beliefs and only partly at work in case of cost beliefs of both boarding-school and basic education-school students, something observed in other regions as well (Arens et al., 2020; Gaspard et al., 2018; Guo et al., 2017; Jansen, Schroeders, et al., 2015; Marsh et al., 2001, 2014; Marsh, Lüdtke, et al., 2015; Möller et al., 2006; Stocker et al., 2021). In the model's most basic version, motivation (e.g., expectancy beliefs) students exhibit in mathematics and language is positively predicted by prior achievement within the matching domain (i.e., hypothesis 2a) and negatively predicted by prior achievement within the non-matching domain (i.e., hypothesis 3a; Möller et al., 2009). These patterns were not observed in case of daily utility across both school types, which confirmed hypothesis 3b that daily utility is not subject to any far-domain effects. We were surprised not to find far-domain effects common especially within the formation of expectancy beliefs of basic-education students. The design of student report cards that are given out to students at both boarding schools and basic-education schools should generally promote dimensional comparisons. School subject-specific school marks within the final report card are presented in form of per centages, which should allow easy comparison of one's own attainment with attainment by others as well as of one's performance across different school domains. At basic-education schools, so it appears, this information may only help to shape cost but not expectancy beliefs. Additional research is required to explore why dimensional comparisons between far domains are not common amongst basic-education students.

Does achievement predict motivation between subjects within the quantitative domain (hypotheses 4a, and 4b)?

Especially later iterations of the GI/ E model proposed assimilation effects of prior achievement especially between near domains that are considered similar (Möller et al., 2016). In case of both basic-education both boarding schools, we each only found 1 out of 6 cross-subject effects to be positive. Hypothesis 4a was therefore not confirmed. Assimilation effects were only observed within cost beliefs. Either mathematics grades predicted Chemistry beliefs (basic-education schools) or vice versa (boarding schools). In line with hypothesis 4b, we did not find any assimilation effects within daily utility for both school types. Hence, whereas the most fundamental GI/ E model logic of within-domain and far-domain effects were observed especially within boarding-school expectancy beliefs, more nuanced model aspects of assimilation effects within near domains of Chemistry and mathematics applied neither to boarding nor basic-education schools. By contrast, using a sample of German secondary students, Jansen et al. (2015) found that between Chemistry and mathematics school grades always predicted expectancy beliefs (self-concept) in the other domain. We are not aware of any study on cross-subject effects on cost beliefs between Chemistry and mathematics. Gaspard et al. (2018) examined respective effects between mathematics and Physics. Similar to the case of cost beliefs at boarding schools, they found Physics achievement to predict to mathematics cost beliefs but not mathematics achievement to predict Physics cost beliefs. Like us, who did not find any assimilation effects amongst daily utility, Gaspard et al. did not any cross-subject effects of prior achievement between mathematics and Physics either.

Does achievement predict motivation between subjects within the language domain (hypotheses 5a, 5b, and 5c)?

The patterns of cross-subject effects within the quantitative domain somewhat resembled those we found within the language domain. In case of boarding schools, we only found one significant cross-path effect of English achievement on Kinyarwanda expectancy beliefs, which contradicted hypothesis 5b of consistent contrast effects between languages. Again, besides the most fundamental GI/E model logic of within-domain and cross-domain effects, more nuanced near-domain effects did not materialize in the language domain at boarding schools. However, as implied by hypothesis 5b the effects of English school grades on Kinyarwanda expectancy beliefs were indeed negative. This may be an indication of the hypothesized multilingualism at boarding schools. In fact, as table 25 highlights boarding school students performed slightly better not only in terms of Kinyarwanda but also English school grades. One can also assume that teachers at boarding schools hold a better command of teaching in English than their colleagues at basic-education schools. According to the reasoning of van der Westhuizen et al. (2022), the slightly higher bilingual environment (i.e., Kinyarwanda as the day-to-day language and English as the main medium of instruction) at boarding schools may allow their students to better detect the subtle differences between English and Kinyarwanda. In fact, even though most cross-subject effects are not statistically significant, their signs often suggest contrast effects of achievement between English and Kinyarwanda across the different motivational beliefs. In case of basic-education schools we found no cross-subject effects amongst value and expectancy beliefs but cost beliefs. Amongst cost beliefs English and Kinyarwanda achievement exerted positive assimilation effects on cost perceptions in the other subject, thereby partly confirming hypothesis 5a. Despite our reasoning, assimilation effects between languages have only been partly reported in the literature (Marsh et al., 2014; Möller et al., 2006). Xu et al., (2013) found no effects between Chinese and English. Other studies reported contrast effects (Arens et al., 2020; Marsh et al., 2001, 2014; Marsh, Lüdtke, et al., 2015; Niepel et al., 2014; Stocker et al., 2021). Hence, unlike at least some educational researchers hypothesized (e.g., Marsh & Yeung, 2001), Rwandan basic-education students may not consider both English and Kinyarwanda as different schools subjects after all. This in turn could be an indication of lower levels of multilingualism at basic-education schools (van der Westhuizen et al., 2022). As Williams (2019) highlights, at basic-education schools both students and teachers often have poor command of the English language. Therefore, English teaching may generally be interspersed with Kinyarwanda (e.g., Honeyman, 2015b; Williams, 2019). As a result, both domains may be perceived to be more similar by basic-education students than it is assumed within motivational theory. Absence of any language-related cross-subject effects on daily utility across both basic-education and boarding schools was in line with hypothesis 5c.

Limitations, Future Research, and Practical Implications

The limitations of our study need to be highlighted. First, given that the study investigated belief formation amongst Rwandan secondary students, a major limitation was the cross-sectional nature of the data collected. To address this shortcoming, replication studies of the GI/ E model logic within the African context are required that employ longitudinal (e.g., Niepel et al., 2014) as well as experimental (e.g., Möller & Köller, 2001) designs. For example, reciprocal effect models allow studying how motivational beliefs and achievement may interact over time (Niepel et al., 2014). Second, amongst all motivational beliefs we found daily utility to be least predicted by the GI/ E model logic. However, these findings may not extrapolate to other value dimensions within the Rwandan context. In fact, other value dimensions such as intrinsic value are well explained by the GI/ E model (e.g., Gaspard et al., 2018; Götz et al., 2008). Future GI/ E model research in Rwanda therefore needs to include other value aspects as well. The study reported by Bulla et al. (2022) can help to operationalize these motivational beliefs within the Rwandan context. Third, we derived the nature of the cross-subject effects (e.g., assimilation effects within the quantitative domain) from the Western literature on the GI/ E model (e.g., Jansen, Schroeders, et al., 2015; Marsh, Lüdtke, et al., 2015; Möller & Marsh, 2013). We did not investigate to what extent Rwandan students actually consider subjects within and across the different domains as similar or different (e.g., Helm et al., 2016). These insights, however, may help to better understand the inner workings of the GI/ E model within the Rwandan context. Fourth, future replications should also control for the socio-economic as well as educational background of the family of students. In fact, family-level factors such as parental attitudes (Harackiewicz et al., 2012a) may constitute an important set of determinants of student motivational beliefs. Controlling for these factors may help to better understand the social and dimensional comparison processes amongst Rwandan students. Fifth, the sample we employed was not nationally representative. To better understand to what extent our findings are indicative of Rwanda as a whole, our study should be replicated utilizing nationally representative samples as well (e.g., van der Westhuizen et al., 2022). Sixth, a key strength of the study may also contribute to its limitation. Rather than importing any of the Western-borne motivational constructs into the Rwandan context, the study relied on motivational measures explicitly developed for this research undertaking in Rwanda through a bottom-up approach. This may have contributed to the study's overall cultural and ecological validity. That way, however, the study may also have contributed to a proliferation of motivational constructs. The reliance on Rwanda-specific motivational constructs may also have hampered the comparison of Rwanda findings especially with those reported in the motivational literature. Within future research, studies on GI/ E models within the African context should also be replicated by combining locally grounded and Western-borne motivational scales.

The practical implications of our study may be at least threefold. First, our study findings highlight that educational achievement and student motivation are related. As our study shows, achievement at least partly predicts motivational beliefs. The correlational nature of our data, however, may also imply that motivational beliefs do predict academic achievement within the Rwandan context as well (e.g., Eccles & Wigfield, 2020). In fact, Bulla et al. (2022) reported a simple listing exercise within their study in which they asked Rwandan students to give examples of those subjects they associate with confidence and motivation. The results at least partly correlated with school grades. Both the results reported here as well as the ones reported by Bulla et al. (2022) therefore justify further investigations into the extent to which student motivation indeed predicts academic achievement of Rwandan students. This in turn may of interest not only to education researchers but also to Rwanda's education practitioners. As is common in other low-income countries as well (e.g., Sandefur, 2018), Rwanda has been struggling with poor learning outcomes amongst its students (e.g., Burdett & James, 2018). Here, motivationally informed school interventions (e.g., Yeager & Walton, 2011) could constitute an additional tool to address learning challenges amongst students in Rwanda and elsewhere. In fact, they may be especially interesting to low-income countries because motivationally informed interventions have been found to be impactful while often not being resource intense. Future research employing experimental designs needs to help further exploring the practical potential of the motivation-achievement nexus (Lazowski & Hulleman, 2016). Second, our study results highlight that also within the Rwandan context educational interventions intended to support student learning may have negative longer-term consequences. Within subjects, any improvement within students' academic achievement most likely also leads to more favourable motivational beliefs. Across domains, however, learning improvements in one domain most likely have negative effects on students' motivational beliefs in other domains. To the extent that motivational beliefs also predict subsequent academic achievement, interventions to improve learning outcomes in one domain would then result in negative effects on learning outcomes in other domains. Hence, our study findings therefore imply that educational interventions to improve learning outcomes in one school area may require auxiliary measures to prevent any contrast effects on other school areas. However, to what extent auxiliary measures are indeed necessarily partly depends on the school context, which is a third implication of our study. The manifestation of motivational (and other learning processes?) seem to differ between boarding and basic-education schools. This might be of interest especially to those African countries[24] that also have school-level ability grouping in the way Rwanda has (Sayeda & Kanjeeb, 2013).

Conclusion

A key insight from our study is that overall the GI/ E model explains motivational beliefs of Rwandan students. This in turn provides further evidence on the applicability of the

[24] Other countries are Botswana, Burkina Faso, Chad, Comoros, the Democratic Republic of Congo, Eritrea, Eswatini, Ethiopia, Kenya, Mali, Mozambique, Niger, Nigeria, Senegal, Sierra Leone, Tanzania, Uganda, Zambia, and Zimbabwe (Sayeda & Kanjeeb, 2013).

model within the African context as well (e.g., Marsh & Hau, 2004). However, the model's applicability appears to be moderated by school context, which we operationalised through the including of basic-education and boarding schools into our sample. On the one hand, the most fundamental logic of the GI/ E model applied fully only to expectancy beliefs reported at boarding schools. In those instances, we found respective motivational beliefs to be shaped by achievement within the matching domain as well as by achievement within far domains. Conversely, achievement within related subjects hardly seemed to shape motivational beliefs neither in the language nor the quantitative domain. In case of basic-education schools, the most basic model logic somewhat applied to cost beliefs only. On the other hand, we found assimilation effects on cost beliefs especially within language domain. In those instances, English and Kinyarwanda achievement positively predicted cost beliefs in the other subject. Amongst boarding schools, we hardly observed any contrast or assimilation effects between similar subjects. These differences in structural relations may suggest that, unlike the person-level factors of age and gender, school context may moderate the unfolding of motivational processes within students. Given that between-school ability grouping is common not only in Rwanda but also in other countries of the continent (Sayeda & Kanjeeb, 2013), systematic motivational research that needs to be promoted in Africa needs to take into account these differences in learning contexts between schools.

5.4. Manuscript three: The generalized internal/external frame of reference model with average motivational states. Findings from Rwanda.

Abstract

The generalized internal/ external frame-of-reference (GI/E) model postulates that prior academic achievement positively predicts motivational beliefs within matching subjects. Across different subjects, prior achievement is postulated to predict motivational beliefs negatively. GI/E model studies generally gauge motivational beliefs through school subject-specific self-reports. In the present study, we tested the GI/E model's applicability to competence and cost beliefs operationalised through average motivational states measured longitudinally. To do so, we used lesson-specific self-reports provided by a sample of 446 Rwandan secondary students (mean age = 15.33 (1.59); female = 57.85%) across 261 English, Kinyarwanda, and mathematics lessons (5214 lesson self-reports in total). GI/E model relations were estimated using latent state/trait models. Within subjects, average English and mathematics competence states and average mathematics cost states were predicted by achievement. Across subjects, only average mathematics competence and cost states were predicted by achievement. The model's partial non-applicability to average stats coincided with the constructs' low level of subject specificity as well as low levels of association with school subject-specific competence and cost self-reports. We took this as evidence that average competence and cost states may be related but distinct constructs from school subject-specific competence and cost self-reports.

Keywords: Generalized internal/external frame of reference (GI/E) model, competence beliefs, cost beliefs, trait, state, Rwanda,

Introduction

As predictors of educational outcomes (e.g., Eccles & Wigfield, 2020), students' beliefs about the level of their own academic abilities (i.e., competence beliefs) and how taxing academic work is (cost beliefs) have been found to vary across school subjects (e.g., Gaspard et al., 2018). According to the generalized internal-external frame of reference (GI/E) model, school subject-specific motivational beliefs are influenced by prior academic achievement. Within a subject, achievement positively predicts competence and cost beliefs. Across subjects that are perceived to be different, achievement in one subject negatively predicts motivational beliefs in the other subject (Möller et al., 2016). Numerous studies have provided empirical support for these model predictions especially in the case of competence beliefs (i.e., academic self-concept; e.g., Möller et al., 2009) but also cost beliefs (e.g., Bulla et al., 2022c; Gaspard et al., 2018). A review of the GI/E model literature reveals that two important aspects. First, only two studies (Bulla et al., 2022b; M.-S. Chiu, 2008) tested the model within the Africa context. Working with Rwandan students, Bulla et al. (2022b) found the GI/E model to partially explain competence and cost beliefs. More empirical tests of the model predictions within the African contexts are warranted to further increase the model's generalizability (Marsh & Hau, 2004). Second, GI/E model studies have mainly been relying on measuring motivational beliefs through school subject-specific self-reports. However, these measures may only partly predict those typical motivational experiences students actually have within a given subject. Such person characteristics are therefore increasingly modelled as the average of momentary motivational states (e.g., Finnigan & Vazire, 2018) that are repeatedly measured within a given subject over a (short) timeframe (intraindividual/ intensive longitudinal design; e.g., Bolger & Laurenceau, 2013). Within the present study, we tested the applicability of the GI/ E model to average competence and cost states inferred for a sample of Rwandan secondary students based on lesson-specific self-reports collected for 13 days. By using a Rwanda sample, we thus contributed to the geographical diversity of GI/E model research (Marsh & Hau, 2004). We argue that the extent to which the model predicts average motivational states is indicated by their level of school subject-specificity (Möller et al., 2016) as well as their level of correlation with those school subject-specific competence and cost self-reports that have been found to be explained the model's logic (Bulla et al., 2022b).

The GI/E model and subject-specific motivational beliefs

According to the GI/E model, school subject-specific motivational beliefs are shaped through the interaction of two self-evaluation processes that each use different reference points. First, students perform social comparisons through evaluating their own academic achievement in one subject by contrasting it with achievements attained by others (external references) in that subject. The higher academic achievement one has attained previously the more favourable one's subsequent motivational beliefs in that subject (Möller et al., 2016). Second, students perform dimensional comparisons through contrasting their own achievement in one school subject with their own achievement in another subject (internal reference). To the extent that both subjects are seen to be similar (e.g., mathematics and Physics), achievement in one subject should positively predict motivational beliefs associated with the other subject (assimilation effects). To the extent that two subjects are different (e.g., mathematics and English), achievement in one subject should negatively predict motivational beliefs in the other subject (contrast effects; Möller et al., 2016; Möller & Marsh, 2013). Motivational beliefs are thus determined through the joint interaction of both dimensional and social comparisons. A multitude of studies have supported the GI/E model predictions especially in the case of self-reported competence (academic self-concept; e.g., Möller et al., 2009) and increasingly cost beliefs (e.g., Bulla et al., 2022c; Gaspard et al., 2018). A review of the GI/E model literature reveals that these studies typically gauge school subject-specific motivational beliefs through self-reports scales that ask respondents to respond to items such as "I am good in [SUBJECT]" (Gaspard et al., 2018, p. 5). It has increasingly been argued that responses to such items are susceptible to respondents' biases and self-schemata and thus may not necessarily indicate what individuals are truly alike (e.g., Finnigan & Vazire, 2018). An alternative approach to gauge a person's stable characteristics relies on state measures. States refer to the behaviours, thoughts, and feelings that are exhibited at a particular time (Baumert et al., 2017). States are rather short-lived phenomena because they constitute interactions between a person's stable characteristics (i.e., traits) and those contextual factors that make up a specific situation (Geiser et al., 2013). The states that individuals report may thus vary from moment to moment. However, stable person characteristics such as school subject-specific motivational beliefs can be modelled as person-specific averages of situation-specific self-reports repeatedly collected within short time frames (intraindividual/ intensive longitudinal design; e.g., Bolger & Laurenceau, 2013). To estimate average states, researchers tend to employ latent state/ trait models, which are guided by the assumption that person's measurement scores observed longitudinally can be decomposed into a person-specific average (i.e., the trait) and time-point specific deviations (i.e., states) from that time-invariant average (e.g., Geiser et al., 2013). Following this logic, a student's school subject-specific motivational beliefs can be modelled as average motivational states that in turn are repeatedly measured within a given subject. We are not aware of any study that applied the GI/E model to average motivational states. A few studies (e.g., Bong, 1998) applied the model to test-related self-efficacy but generally found its logic not to apply to these situational measures (see discussion in Möller et al., 2009).

School subject-specificity and convergent validity as first indications of the GI/E model applicability

The applicability of the GI/E model to a motivational construct may be foreshadowed by the following conditions. First, the constructs needs to be subject specific in the sense that it is weakly or even negatively correlated across different school subjects (Möller et al., 2016). Lack of school subject-specificity may indicate poor applicability of the model to a construct. Using a second-order factor model, Bong (1998) inferred domain-specific competence beliefs based on test-related self-efficacy items, an approach loosely comparable to latent state/ trait models. Across the mathematics and verbal domain, competence was substantially correlated (.70), which indicates low school subject-specificity. Moreover, especially verbal competence beliefs were not predicted by achievement across both domains. Second, motivational constructs that each are well explained by the GI/E model appear to be substantially correlated, something we interpret as some form of convergent validity. According to Gaspard et al. (2018), for example, academic self-concept was well explained by the GI/E model whereas social utility was not (both were measured using school subject-specific self-reports). Social utility did not only exhibit low levels of school subject-specificity but was also only moderately correlated (.32) with academic self-concept. By contrast, constructs such as self-concept and intrinsic value that were all predicted by the GI/E model exhibited strong correlations between.74 and .85 in absolute terms. Hence, if we use Bong's (1998) finding as a reference, we would expect that average motivational states do not only exhibit low subject specificity but also weak to moderate convergent validity vis-à-vis those school subject-specific motivational beliefs that have already been explained by the GI/E model. As the next section shows, students' school subject-specific competence and cost self-reports in turn have been explained by the GI/E model.

The Rwandan School Context

In the East-African country Rwanda, Kinyarwanda is the national language (NISR, 2016). Students have access to 12 years of free and basic education (12YBE). After six years of primary schooling, thy attend three years of lower (S1 – S3) and then upper secondary education (S4 – S6). Traditionally, boarding schools were the providers of secondary education and are still seen as "schools of excellence" (Williams, 2019, p. 648). In the wake of 12YBE, the Rwandan government introduced basic-education schools. Unlike boarding schools, attendance at basic-education schools is not tied to any substantial tuition fees nor prior academic performance during primary years. Because of the latter, basic-education schools are often considered to be "schools for those who failed" (Williams, 2019, p. 646). Using school subject-specific self-reports, Bulla et al. (2022b) examined the applicability of the GI/E model to students' motivational beliefs levels associated with Chemistry, English, Kinyarwanda, and mathematics amongst both basic-education with boarding schools. In case of boarding schools, the model explained both competence (i.e., expectancy) and cost (an omnibus variable combining emotional costs and effort) beliefs whereas amongst basic-education schools it mainly predicted cost beliefs. In each case, cross-subject correlations were also weak to moderate. Hence, convergent validity of average competence and cost states vis-à-vis self-reported competence and cost beliefs may thus be yardstick by which to predict the applicability of the GI/E model to the former two constructs. Bulla et al. (2022b) also found age and gender to partly predict students' motivational beliefs. Older students tended to have less favourable cost beliefs especially within the language domain. Amongst basic-education students, girls tended to exhibit less favourable cost and competence beliefs than boys. Hence, GI/E studies within the Rwandan context best control for age, gender, and school type as well.

The present study

In the present study, we examined the applicability of the GI/E model to Rwandan students' average competence and cost states longitudinally gauged at the lesson-specific level within the subjects of English, Kinyarwanda, and mathematics. The extent to which this holds true, so we argue, can be predicted by the average-state constructs' level of specificity as well as level of correlations with other school subject-specific motivational beliefs. The present investigation was therefore guided by the following hypotheses developed based on the above literature review:

1) We assumed that average competence and cost states exhibit low levels of subject specificity. In line with Bong (1998), we expected that both average competence and cost states exhibit substantial cross-subject correlations.
2) We also expected only moderate convergent validity between average competence and cost states on the one side and school subject-specific competence and cost self-reports on the other side. To be more precise, we expected respective correlations to be substantially below .75 (Gaspard et al., 2018).
3) In line with the expected low subject specificity and convergent validity, we predicted the GI/E model to only partly explain average competence and cost states. We did not expect the internal frame-of-reference effects to apply to the average competence and cost states (Bong, 1998) manifested through regression coefficients that are not significantly different from zero.

4) Even though the GI/E model may not predict well average competence and cost states, we expected that average competence and cost states are linked to prior school grades.
5) We expected different covariates to predict average competence and cost states (Bulla et al., 2022b). To be precise, we expected (5a) school type (i.e., basic education versus boarding schools), (5b) gender, and (5c) age to influence the levels of average competence and cost states.

Method

Procedures

The study reported here was part of a wider project on achievement motivation in Rwanda carried out between January and February 2020. Data collection started with a cross-sectional survey of students' competence and cost beliefs in English, Kinyarwanda, and mathematics (Bulla et al., 2022b). In the subsequent three weeks, self-reports on lesson-specific competence and cost states were collected from students within English, Kinyarwanda, and mathematics. In order not to disrupt teaching and learning, self-reports were only collected after teachers ended lessons (i.e., event-contigent design; e.g., Bolger & Laurenceau, 2013). The questionnaire, translated into Kinyarwanda, always collected information on personal information (i.e., name, age, gender, class), which was then used to track student responses across lessons. On average, it took around five minutes for a class to finish questionnaire completion. Within the present study, we report the data collected across 261 lessons (41.0% in mathematics, 37.2% in English, and 21.8% in Kinyarwanda). Table 27 on the next page provides a breakdown of the lessons covered by school as well as subject. It also highlights that data on Kinyarwanda was hardly collected at boarding school B2. We were only able to include five Kinyarwanda lessons from that school. Across 261 lessons covered, 5214 lesson self-reports were collected (1879 in English, 1171 in Kinyarwanda, and 2164 in mathematics). The average number of self-reports provided by students was 11.69 ($SD = 3.02$) across all three subjects with 4.21 ($SD = 1.64$) in English, 2.63 ($SD = 1.26$) in Kinyarwanda, and 4.85 ($SD = 2.01$) in mathematics. For students to partake in the study, they were asked to sign a child assent form. Head-teacher consent was sought as they acted *in loco parentis* (British Psychological Society, 1992). As a token of appreciation, all students that submitted consent forms participated in a tombola carried out after field work. Per class, three students won a pocket calculator worth around 75p. The study has ethics approval from University of Oxford (ED-CIA-19-084) and University of Rwanda (266/CMHS IRB/2018).

Table 27: *Distribution of lessons attended during data collection (manuscript three)*

School	English	Kinyarwanda	Mathematics	Total
A2	23	14	32	69
B1	58	38	53	149
B2	16	5	22	43
Total	97	57	107	261

Participants

In the present study, we report data provided by 446 students attending 20 different classes across three public schools inside and outside Kigali. Two out of those three schools were basic education schools (A2 and B1). The last one (B2) was a boarding school. The average number of students we used per class was 22.3 ($SD = 12.54$) and ranged from one to 51 students. As table 28 shows, students from basic-education schools tended to be older and more females than the ones from the boarding school (B2).

Table 28: Descriptive school statistics (manuscript three)

		A2	B1	B2	Average
Boarding school?		No	No	Yes	
N (% female)		154 (57.14)	217 (60.83)	75 (50.67)	446 (57.85)[1]
Age (se)		15.42 (1.50)	15.63 (1.65)	14.26 (1.07)	15.33 (1.59)
		English			
Average score: belief scales	Competence	2.23 (.43)	2.26 (.45)	2.63 (.40)	2.34 (.46)
	Cost	2.15 (.51)	2.12 (.49)	2.20 (.53)	2.16 (.51)
Average score: state scale	Competence	2.32 (.51)	2.57 (.50)	2.74 (.44)	2.51 (.52)
	Cost	2.22 (.52)	2.31 (.57)	2.53 (.55)	2.31 (.56)
Average School grades		55.17 (7.67)	51.61 (17.52)	54.28 (7.16)	53.30 (13.41)
		Kinyarwanda			
Average score: belief scales	Competence	2.56 (.44)	2.67 (.36)	2.77 (.34)	2.65 (.39)
	Cost	2.33 (.54)	2.34 (.55)	2.29 (.61)	2.33 (.55)
Average score: state scale	Competence	2.48 (.49)	2.70 (.45)	2.63 (.48)	2.62 (.48)
	Cost	2.24 (.54)	2.37 (.61)	2.39 (.56)	2.33 (.59)
Average School grades		37.19 (12.57)	66.10 (10.43)	66.61 (7.62)	56.18 (17.56)
		Mathematics			
Average score: belief scales	Competence	2.12 (.44)	2.09 (.49)	2.33 (.44)	2.14 (.47)
	Cost	2.06 (.47)	1.87 (.47)	1.90 (.50)	1.95 (.48)
Average score: state scale	Competence	2.41 (.53)	2.41 (.56)	2.53 (.48)	2.43 (.53)
	Cost	2.25 (.53)	2.19 (.57)	2.26 (.55)	2.23 (.55)
Average School grades		38.99 (13.47)	38.32 (12.79)	52.79 (13.70)	40.98 (14.19)

Note.[1]) this statistic refers to the total and not average. ‘N’ stands for sample size. The numbers in brackets refer to standard errors associated with the average scores. ‘se’ refers to standard errors.

Measures

To test for convergent validity between the school subject-specific self-reports and the average states, we wanted to make sure both types of scales are as similar as possible (see table 29). Both sets of scales were exclusively developed for the Rwanda project this study was part of. Bulla et al. (2022b) outlined the process followed for the development of the school subject-specific competence and cost beliefs. In a similar vein, we developed lesson-specific scales. Lesson-specific competence and cost were each measured through three items. To measure self-reported competence and cost beliefs, we used the scales developed by Bulla et al. (2022b) but adapted them in case of the cost beliefs. Both the school subject-specific and the lesson-specific scale on competence were similar in that each of their items were cognitive in nature (e.g., did you understand it). The initial scale on school subject-specific cost beliefs also contained emotion items, which we removed. It shortened the scale to three items as well but was then similar to the lesson-specific cost scale in terms of the items it used. All motivational scales had a three-options answer format: 1) often; 2) sometimes; or 3) never and were coded in such a way that higher value reflect more favourable beliefs.

Table 29: *The motivational scales to infer lesson-specific states and subject-specific beliefs*

Item	Competence	Item	Cost
Self-reported motivational belief scale			
BC1	Do you understand [SUBJECT]	BE1	Is the subject of [subject] difficult?
BC2	Do you do well in [SUBJECT]?	BE2	Is the subject of [subject] hard?
BC3	Do you know the [SUBJECT]?	BE3	Is the subject of [subject] tiring?
BC4	Do you get good marks in [SUBJECT]?		
BC5	Are you capable of [SUBJECT]?		
BC6	Do you understand [SUBJECT]?		
Self-reported motivational state scale			
LC1	Do you think you understood today's activities of [SUBJECT]?	LE1	Were today's activities of [SUBJECT] hard?
LC2	Do you think you were capable to do today's activities of [SUBJECT]?	LE2	Were today's activities of [SUBJECT] tiring?
LC3	Do you think you understood today's activities of [SUBJECT]?	LE3	Were today's activities of [SUBJECT] confusing?

For each scale, table 30 presents average polychoric correlations across items, which we interpreted as measures of reliability. Amongst school subject-specific scales, the average correlation was .64. Amongst lesson-specific scales, the average correlation was .58 at the at the within level and .95 at the between level. We considered these correlations as substantial and thus concluded the scales as reliable. Table 28 also presents average scores for both school subject-specific self-reported competence and cost beliefs and lesson-specific competence and cost self-reports. Prior academic achievement was operationalised through school grades (e.g., Möller et al., 2009) from the 2019 school year (table 28). In Rwanda, they are expressed in terms of per centages and represent the averages students attain across the three school terms of the Rwandan school year. School exams within the third term are prepared by the national level and uniformly administered across schools (Nzabalirwa, 2014). Hence, to some extent end-of-year school marks are comparable across schools. For the analysis, we standardised school grades at the school level, it helped the models converge.

Table 30: Average within-scale polychoric correlations (manuscript three)

	English	Kinyarwanda	Mathematics	Average
Average single-level polychoric correlations (domain-specific self-reports)				
Competence	0.74	0.72	0.74	0.73
Cost	0.58	0.64	0.44	0.55
Average	0.66	0.68	0.59	
average within-level polychoric correlations (lesson-specific self-reports)				
Competence	0.62	0.59	0.67	0.63
Cost	0.55	0.51	0.55	0.54
Average	0.59	0.55	0.61	
average between-level polychoric correlations (person-specific self-reports)				
Competence	0.96	0.91	0.98	0.95
Cost	0.97	0.99	0.94	0.96
Average	0.97	0.94	0.96	

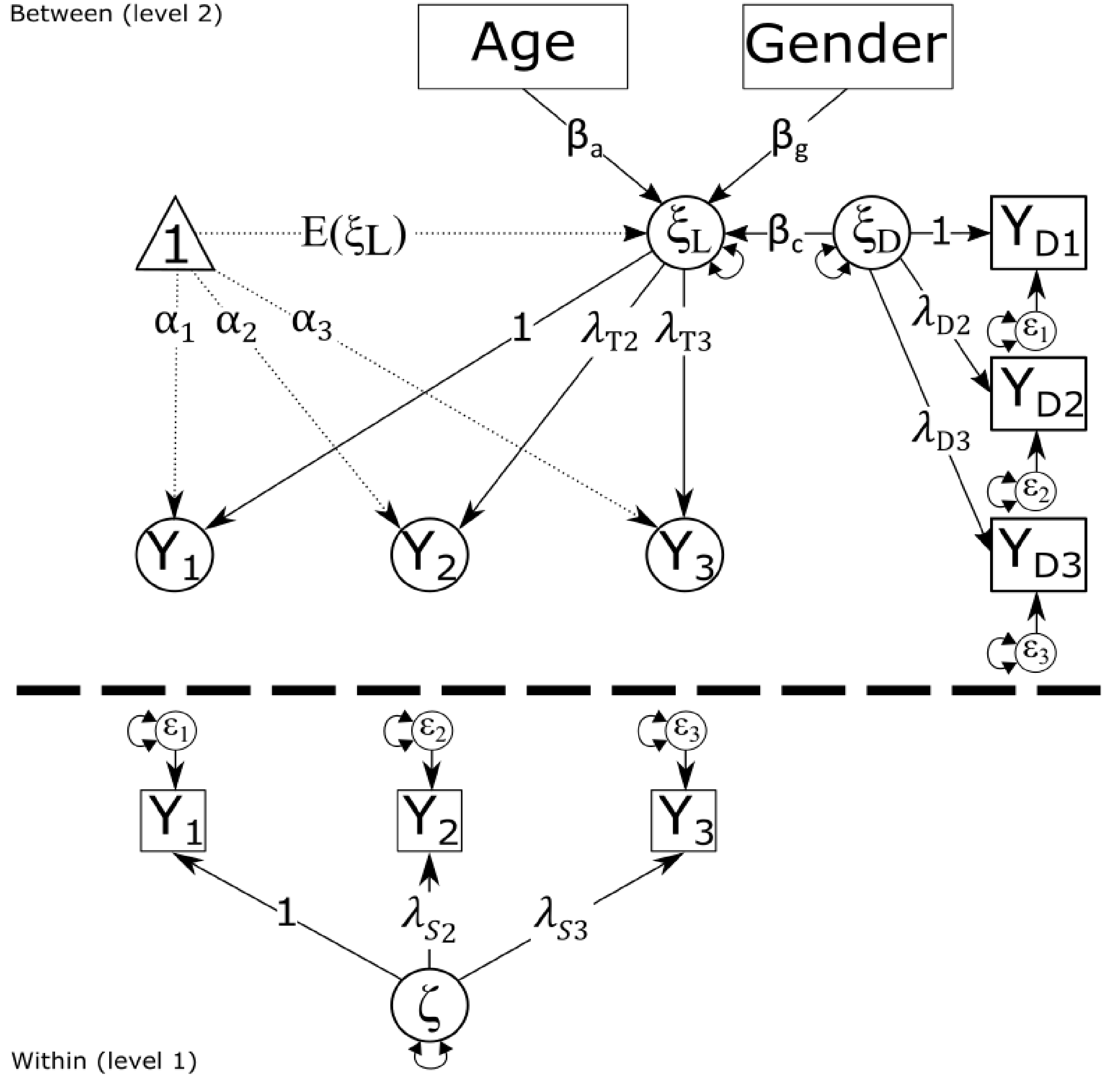
Note. Polychoric correlations were estimated using a multilevel framework. Level one represented the lesson-specific within-person level and was nested within level two which constituted the between-person level.

Latent Single Trait Multiple States (STMS) Modelling

The data was hierarchical in the sense that lesson-specific self-reports were nested within students. The data was also imbalanced because the number of lesson-specific self-reports varied across respondents and subjects. We therefore analysed the data using a multiple-indicator parameterization of the latent single-trait multiple state (STMS) model (Geiser et al., 2013). Within STMS models, person-specific averages (i.e., the traits) are located at level two (i.e., the person-specific between level) while time-point specific deviations (i.e., states) being located at level one (i.e., the situational within level). State residuals as well as error variables were implied to have zero means and to be uncorrelated across time. State residuals were specified to be uncorrelated with latent traits. Model intercepts and factor loadings were all assumed to be time invariant (i.e., “pure state-variability model”; Geiser et al., 2013, p. 5). In line with the assumption of cross-level measurement invariance (Stapleton et al., 2016), factor actor loadings were also constrained to be equal between levels. For both states and traits, the first factor loading was fixed to one. To ensure fast model convergence, we estimated models separately for each construct type (competence versus cost).

To test hypothesis 2, we estimated a STMS model for each construct with the corresponding school subject-specific motivational beliefs level-two predictor (β_c) as well as age (β_a), gender (β_g) and school dummies as controls (see figure 14 on the next page). Across the six models, missingness ranged from 1.67% (cost beliefs in English model) to 9.22% (competence beliefs in Kinyarwanda model). To investigate the remaining sets of hypotheses, we always combined two school subject-specific STMS setup not one model. In this setup, school subject-specific state items were arranged column wise with lesson-specific self-report for each item row-wise. Given the imbalanced nature of the data, we obtained substantial missingness within each model (see table 31). The level of missingness would have been even higher, had we tried to run a model across all three subjects. School grades, age, gender, and school dummies were included as level-two variables. The joint model was always jointly estimated for two of the three subjects (see figure 14). Again, we run this series of models for each motivational construct separately resulting in 3 (i.e., English/ Kinyarwanda, English/ mathematics, and Kinyarwanda/ mathematics) x 2 (i.e., competence and cost) = 6 models. Within each subject combination, we could then estimate correlations between average states (ρ_{ξ}) and school grades (ρ_S). We could then also estimate within-subject (β_{11} and β_{22}) and cross-subject effects (β_{12} and β_{21}). Since each subject was always included in two of the three models, we also obtained two estimates for each within-subject effect as well as covariate effects (see table 34 on page 243). Besides one exception[25], however, the two estimates for a parameter did not substantially vary across models. Because the motivational data was ordinal in nature, we used Mplus' CATEGORIAL command within all model estimations.

Figure 14: The hypothesized structural model to estimate convergent validity (manuscript three)



Note. (based on Geiser et al., 2013) used to estimate correlations (convergent validity) between average states and corresponding school subject-specific self-reported beliefs. The latter were modelled based on three (cost beliefs) or six items (competence beliefs).

---

[25] In two cases, one of the school dummies turned out to be significantly different from zero in one model but insignificantly different from zero in the related model. Since the results were inconsistent, we interpreted the overall effect of this school dummy as insignificant across both models.

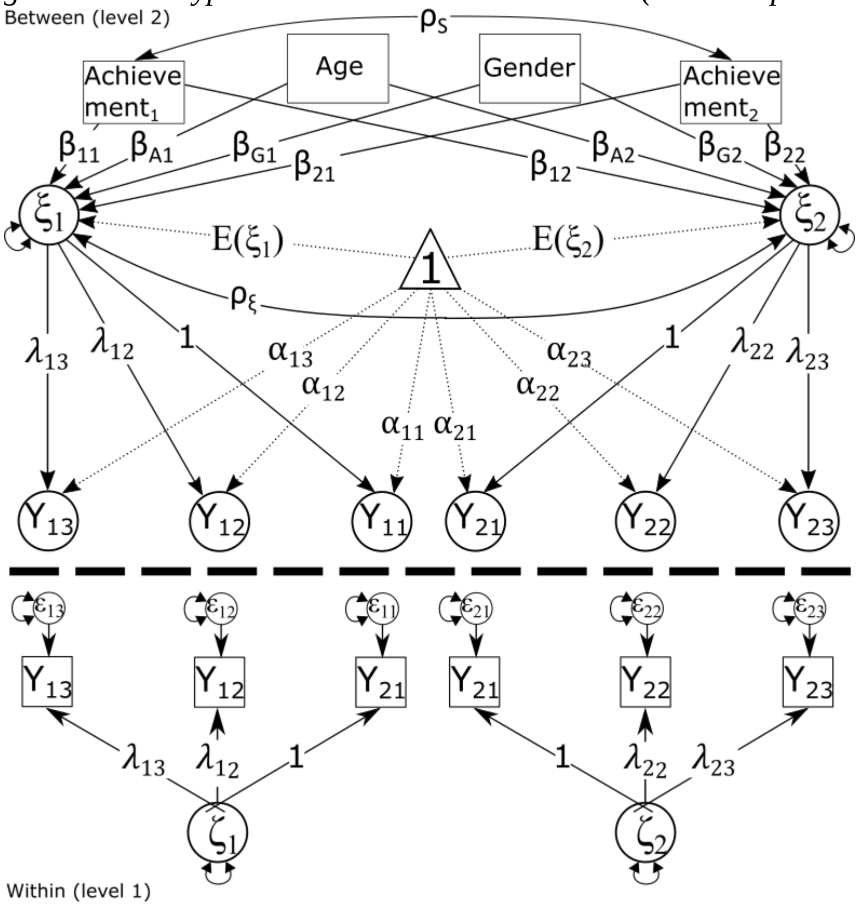
Table 31: Missingness within datasets (manuscript three)

Model	Competence in subject = 1	Competence in subject = 2	Cost in subject = 1	Cost in subject = 2
1 = English & 2 = Kinyarwanda	29.80%	55.76%	29.41%	56.41%
1 = English & 2 = Mathematics	43.37%	35.41%	43.05%	35.49%
1 = Kinyarwanda & 2 = Mathematics	58.26%	23.75%	58.88%	23.83%

Bayesian Estimation

All models were estimated through a Bayesian approach using Mplus version 8.7 (L. K. Muthén & Muthén, 2017) and a seed value of 21012020. Bayesian estimation was chosen because running multilevel latent models with categorical data requires large numbers of dimensions of integration, which often comes with low numerical precision. The posterior distribution was obtained using the iterative Markov chain Monte Carlo (MCMC) algorithm (van de Schoot et al., 2014) with the default option of two chains with two processors used for parallel computing. Using the potential scale reduction (PSR) factor, we specified the MCMC convergence criterion to be 0.01 (Hox et al., 2012), which is more robust than Mplus' default setting of 0.05 (van de Schoot et al., 2014). During the computation of the final posterior distribution, only every 30th iteration was used (thinning), which reduced the effects of parameter autocorrelation between subsequent chain iterations. To ensure convergence, we discarded the first 5000 iterations as 'burn-in'. To check model convergence, we rerun each model again but this time with the number of iterations required being twice the number of iterations the corresponding model needed in the first place. To do so, we used Mplus' FBITERATIONS option. During all model estimations, we used Mplus' diffuse priors (for their hyperparameters see L. K. Muthén & Muthén, 2017) because no intrapersonal motivational research was carried out in Rwanda that would have allowed deriving any informative priors. Given that our sample size was rather large, we expected the use of the default priors as less consequential. To perform convergence diagnostics, we inspected the trace plots. We checked whether or not all chains converged to the same target distribution and whether or not all iterations were based on stable chains. Model fit itself was assessed through the posterior predictive p-value (PPP), which are computed using the likelihood-ratio chi-square statistic (Gelman, 2014). Posterior predictive p-values (PPP) of 0.5 as well as a confidence interval for the chi-square statistic that contains the zero value are taken as evidence of excellent model fit. With the sole exception of the model on convergent validity for English competency, model fits generally turned out to be acceptable (PPP-values between .225 and .590, see table 32 through 34) with all 95% confidence intervals containing the zero value.

Figure 15: The hypothesized structural GI/E model (manuscript three)



Note. (based on Geiser et al., 2013) used to estimate structural relations between prior school grades and subsequent average states. Within each model, we also controlled for school dummies (not depicted here).

Results and discussion

Table 32 depicts means for both direct and latent trait beliefs. For latent trait beliefs, table 32 also depicts level-1 and level-2 variances and the corresponding intraclass-correlations (ICCs). As the ICCs highlight, within-student variances were relatively substantial. Below, the results of tests of each of the five sets of hypotheses are presented.

Average competence and cost states are substantially correlated across school subjects (hypothesis 1)

Cross-subject correlations (ρ_ξ in figure 15) for each of the average states are presented in row 'Trait1 - Trait2' in table 34. All correlations were substantially different from zero and exhibited strong effects ($\bar{\rho}_{\xi all} \approx .71$). We therefore considered hypothesis 1 on the low school subject-specificity of both constructs as supported by the data. Cross-subject correlations were slightly stronger in the case of average competence ($\bar{\rho}_{\xi call} \approx .68$) than average cost ($\bar{\rho}_{\xi eall} \approx .74$) states. Cross-subject correlations appeared to be stronger across Kinyarwanda and mathematics ($\bar{\rho}_{\xi km} \approx .723$) and somewhat smaller across English and mathematics ($\bar{\rho}_{\xi em} \approx .69$). Given school subject-specificity being a model prerequisite (Möller et al., 2016), we took these correlational results as a first indication of the GI/E model most likely not to predict average competence and cost states well (Bong, 1998; Bulla et al., 2022b; Gaspard et al., 2018). Row 'Ach1 - Ach2' in table 34 also depicts cross-subject correlations (ρ_s in figure 15) of school grades. With an average of just around .15, school grades appears to be significantly less correlated across subjects than what is normally reported withing the GI/E model literature (e.g., Möller et al., 2009). The majority (83%) of our sample was from basic-education schools, a student population for which Bulla et al. (2022b), too, found small average correlations (.23) within school grades. Respective corelations were significantly higher (.45) amongst boarding schools. The weak correlation we observed thus may have been due to the idiosyncrasies of our sample.

Rwandan students' average competence and cost states are distinct constructs from self-reported school subject-specific competence and cost beliefs (hypothesis 2)

Table 33 presents the standardised regression coefficients (β_c in figure 14) for the path from self-reported beliefs to average states controlled for school, age, and gender effects. The effects were all substantially different from zero in case of both competence ($\bar{\beta}_{ccom} \approx$ .45) and cost ($\bar{\beta}_{ceff} \approx$ .42).[26] In terms of strength, the effect sizes were comparable to ones reported in the literature. Malmberg et al. (2013) estimated the correlation between self-reported competence (i.e., ability) beliefs and the average of a corresponding state measure to be .49. Ahmed et al. (2010) estimated the correlation between competence self-beliefs and average competence states to be .44. In case of cost, however, our results slightly differed from the ones reported by Malmberg et al. (2013). They estimated the correlation between person-specific difficulty perceptions and the average of task-specific difficulty perceptions to be .21. However, the correlations we observed were far below .75, which was observed between constructs that were otherwise explained by the GI/E model (Gaspard et al., 2018). In line with hypothesis 2, we therefore concluded moderate convergent validity for both average competence and cost states vis-à-vis competence and cost beliefs. Due to moderate

[26] We also run models on zero-order correlations between direct and latent trait beliefs. Effect sizes were somewhat comparable to the regression results.

convergent validity, we also concluded that at least in the case of Rwandan students' average competence and cost states measure different constructs than the ones operationalised through school subject-specific competence and cost beliefs. Given weak subject specificity and moderate convergent validity, we also expected the GI/E model not to explain well average competence and cost states exhibited by Rwandan students. As table 33 highlights, effects were somewhat smaller in the case of Kinyarwanda competence ($\beta_{ckc} \approx .38$) and mathematics cost ($\beta_{cme} \approx .38$). Age, gender, and school hardly exerted any effects. Only in the case of Kinyarwanda competence and mathematics cost, we found some small negative effects of age. Older students exhibited less favourable cost beliefs. School effects only seemed to matter within English. Students from basic-education schools seemed to have exhibited less favourable competence ($\bar{\beta}_{sec} \approx -.25$) and cost beliefs ($\bar{\beta}_{see} \approx -.33$).

Table 32: Descriptive scale statistics (manuscript three)

Item	Direct trait measures	Latent trait measures			
	Mean (standardised)	Intercept_{bet} (standardised)	Variance Level 1	Variance Level 2	ICC
English					
Competence	**.698 [.532, .871]**	1.394 [-.023, 2.802]	3.751	3.529	0.484
Cost	**.321 [.093, .514]**	.612 [-.949, 2.109]	2.124	2.556	0.547
Kinyarwanda					
Competence	**1.377 [1.186, 1.579]**	**2.064 [.629, 3.562]**	4.160	4.936	0.542
Cost	**.613 [.431, .826]**	**1.800 [.402, 3.258]**	2.333	3.154	0.574
Mathematics					
Competence	**1.377 [1.186, 1.579]**	**2.064 [.629, 1.579]**	3.913	2.963	0.430
Cost	**-.259 [-.460, -.066]**	**1.912 [.186, 3.563]**	1.892	1.885	0.499

Note. Level one = lesson-specific level; level two = student-specific level; ICC = intraclass correlation. Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits: competence within English: Sample size = 445; PPP = 0.030, f (95% CI) = -1.79 - 94.31; cost within English: Sample size = 445; PPP = 0.295, f (95% CI) = -25.16 - 46.77; competence within Kinyarwanda: Sample size = 445; PPP = 0.269, f (95% CI) = -33.00 - 64.21; cost within Kinyarwanda: Sample size = 445; PPP = 0.505, f (95% CI) = -34.45 - 34.33; competence within Mathematics: Sample size = 445; PPP = 0.225, f (95% CI) = -29.47 - 66.40; cost within Mathematics: Sample size = 445; PPP = 0.233, f (95% CI) = -24.89 - 48.05. All models control for schools. The coefficients for these control variables are not reported here.

Average competence and cost states are not explained by internal frame-of-reference effects (hypothesis 3)

The rows 'Ach1 – Trait2' and 'Ach2 - Trait1' in table 34 present the standardised coefficients for the within-subject effects of school grades (see β_{12} and β_{21} in figure 15). Only in the case of English (β_{cem} ≈ .143 [.043, .242]) and Kinyarwanda (β_{ckm} ≈ .105 [.005, .203]) on average mathematics competence states and English achievement on average mathematics cost states (β_{eem} ≈ .114 [.013, .212]) did we find significant cross-subject effects. Hypothesis 3 supported in the cases of English and Kinyarwanda but refuted in the case of mathematics. All cross-subject effects were positive. According to the dimensional-comparison theory (Möller & Marsh, 2013), this would suggest that respective subjects are perceived to be similar. The assimilation effects we found were in contrast to Bong (1998), whose study design we considered related to ours. Students in her sample, however, might have been taught mainly in their national language. In Rwanda, however, the medium of instruction (English) is different to the national language (Kinyarwanda). In those contexts, lesson-specific competence and cost (self-)evaluation may not only be about perceived ability levels and perceived difficulty associated with the learning tasks studied in class. They may also be about perceived ability levels and levels of difficulty associated with the language the teacher uses to guide these tasks. The latter aspect may hold especially for basic-education students who have generally poor command over English (Williams, 2019). Students with higher school grades may exhibit higher command over English and thus may not only feel more confident about but also find it easier to follow teaching in English, which in turn may lead to more positive mathematics-related competence and cost states on average. Due to low English levels of both students and teachers, teachers often use a hybrid of both English and

Kinyarwanda in class (Williams, 2019). Hence, (perceived) Kinyarwanda communication skills, proxied through Kinyarwanda school grades, may thus shape mathematics-related competence and cost states, too.

Table 33: Correlations between direct trait and latent trait measures (manuscript three)

Predictor	Latent trait belief	
	Competence	Cost
	English	
Corresponding trait measure	**.488 [.382, .585]**	**.435 [.320, .542]**
Age	.001 [-.102, .107]	.051 [-.052, .157]
Gender	-.001 [-.201, .200]	.027 [-.074, .126]
	Kinyarwanda	
Corresponding trait measure	**.379 [.255, .496]**	**.436 [.316, .547]**
Age	**-.133 [-.247, -.018]**	-.112 [-.222, -.001]
Gender	.013 [-.099, .124]	-.014 [-.121, .094]
	Mathematics	
Corresponding trait measure	**.493 [.390 .586]**	**.383 [.268, .489]**
Age	-.119 [-.227, -.010]	**-.114 [-.225, -.001]**
Gender	-.023 [-.123, .080]	-.032 [-.136, .074]

Note. Level one = lesson-specific level; level two = student-specific level; ICC = intraclass correlation. Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits: competence within English: Sample size = 445; PPP = 0.030, f (95% CI) = -1.79 - 94.31; cost within English: Sample size = 445; PPP = 0.295, f (95% CI) = -25.16 - 46.77; competence within Kinyarwanda: Sample size = 445; PPP = 0.269, f (95% CI) = -33.00 - 64.21; cost within Kinyarwanda: Sample size = 445; PPP = 0.505, f (95% CI) = -34.45 - 34.33; competence within Mathematics: Sample size = 445; PPP = 0.225, f (95% CI) = -29.47 - 66.40; cost within Mathematics: Sample size = 445; PPP = 0.233, f (95% CI) = -24.89 - 48.05. All models control for schools. The coefficients for these control variables are not reported here.

Average competence and cost states are predicted by prior school grades in the matching school subjects (hypothesis 4)

In table 34, the rows on 'Ach1 – Trait1' and 'Ach2 – Trait2' present the within-subject effects associated with prior school grades (see β_{11} and β_{22} in figure 15). In the case of English competence (β_{cee} ≈ .114 [.010, .215] /.127 [.016, .240]) as well as mathematics competence (β_{cmm} ≈ .309 [.207, .405]/ .318 [.216, .433]) and cost states (β_{emm} ≈ .253 [.147, .352]/ .262 [.158, .361]), we found significant effects. As for Kinyarwanda, we considered hypothesis 4 entirely unsupported. In case of cost, we considered hypothesis 4 mainly unsupported whereas in case of average competence states, we considered it largely supported. In case of mathematics, we considered hypothesis 4 fully supported. As per the GI/E model, within-subject effects of prior school grades on average states may suggest the latter constructs are shaped by social-comparison processes. Unlike in the case of test-related self-efficacy items (e.g., Bong, 1998), where respondents are provided with an explicit criterium by which to judge their competence levels (e.g., "how confident are you that you can do this task?), our lesson-specific measures asked students to indicate whether or not they could follow a lesson without providing any reference points. This may indeed encourage social comparisons. However, we administered all lesson-specific scales within the first five minutes after a lesson was concluded. This might not have given students sufficient time to compare themselves with others in class in terms of perceived competence and cost levels. It therefore might have been possible that students indeed relied on other sources of those evaluative information required to respond to our lesson-specific items. One source could have been prior academic successes and failures (i.e., mastery experiences; e.g., Usher & Pajares, 2009). They have been found to shape at least competence states (i.e., self-efficacy) and thus may also shape cost states. Prior mastery experiences in turn may correlate with prior school grades, which in turn may explain why school grades favourably predict competence and cost states. Hence, we may indeed find prior academic achievement to predict average competence and cost states within and between subjects. However, it may do so for reasons other than the ones postulates by the GI/E model. Future research needs to further explore this idea.

Table 34: Estimated coefficients of the structural models (manuscript three)

Predictor	Model English/ Kinyarwanda: 1 = English and 2 = Kinyarwanda		Model English/ Mathematics: 1 = English and 2 = Mathematics		Model Kinyarwanda / Mathematics: 1 = Kinyarwanda and 2 = Mathematics	
	Trait = Comp.	*Trait = Cost*	*Trait = Comp.*	*Trait = Cost*	*Trait = Comp.*	*Trait = Cost*
			Factor correlations			
Ach1 - Ach2		**.408 [.326, .483]**		.085 [-.008, .176]		-.057 [-.149, .037]
Trait1 - Trait2	**.634 [.510, .746]**	**.801 [.717, .874]**	**.675 [.568, .766]**	**.701 [.608, .779]**	**.723 [.603, .828]**	**.723 [.621, .811]**
			Path coefficients			
Ach1 - Trait1	**.127 [.016, .240]**	.104 [-.011, .217]	**.114 [.010, .215]**	.090 [-.017, .192]	.077 [-.027, .181]	.014 [-.094, .118]
Ach2 – Trait1	-.019 [-.126, .089]	.000 [-.110, .110]	-.010 [-.111, .091]	.074 [-.027, .174]	.035 [-.078, .147]	.077 [-.034, .185]
Ach1 – Trait2	.080 [-.040, .198]	.052 [-.066, .169]	**.143 [.043, .242]**	**.114 [.013, .212]**	**.105 [.005, .203]**	.050 [-.050, .151]
Ach2 - Trait2	.041 [-.078, .159]	-.015 [-.132, .102]	**.309 [.207, .405]**	**.253 [.147, .352]**	**.318 [.216, .413]**	**.262 [.158, .361]**
			Covariates			
Age – Trait1	.005 [-.116, .127]	.013 [-.109, .136]	.003 [-.116, .118]	.026 [-.092, .141]	**-.168 [-.291, -.036]**	**-.155 [-.275, -.026]**
Age – Trait2	**-.174 [-.302, -.040]**	**-.157 [-.283, -.027]**	-.109 [-.227, .009]	-.111 [-.228, .009]	-.114 [-.234, .008]	-.123 [-.243, .002]
Gender – Trait1	-.050 [-.151, .053]	-.014 [-.117, -.090]	-.049 [-.152, .057]	-.015 [-.118, .090]	-.025 [-.140, .090]	-.052 [-.166, .061]
Gender – Trait2	-.015 [-.131, .101]	-.031 [-.142, .080]	-.013 [-.118, .094]	-.006 [-.112, .100]	-.037 [-.141, .068]	-.029 [-.135, .076]

Note. Each model was also controlled for school effects (boarding school B2 was the reference). However, these coefficients are not reported here. Model fits: model English/ Kinyarwanda (competence): PPP = 0.575, f (95% CI) = -43.97 - 33.45; model English/ Kinyarwanda (cost): PPP = 0.542, f (95% CI) = -41.385 - 37.55; model English/ Mathematics (competence): PPP = 0.535, f (95% CI) = -41.58 - 38.59; model English/ Mathematics (cost): PPP = 0.517, f (95% CI) = -40.48 - 39.63; model Kinyarwanda/ Mathematics (competence): PPP = 0.590, f (95% CI) = -45.73 - 33.94; model Kinyarwanda/ Mathematics (cost): PPP = 0.524, f (95% CI) = -43.46 - 39.17. Comp. = competence; Ach1 = achievement in school subject 1; Ach2 = achievement in school subject 1; Trait1 = trait in school subject 1; Trait2 = trait in school subject 1.

Average competence and cost states are predicted by gender, age, and school effects (hypotheses 5a through 5c)

Table 34 also presents the two model-specific effects associated with age and gender (also see β_{A1}, β_{A2}, β_{G1} and β_{G2} in figure 15). In the case of gender, we did not find any effects significantly different from zero. We therefore refuted hypothesis 5b. In case of age, we only found small but significant effects on average Kinyarwanda competence ($\beta_{cak} \approx -.168$ [-.291, -.036]/ -.174 [-.302, -.040]) and cost states ($\beta_{eak} \approx -.155$ [-.275, -.026]/ -.157 [-.283, -.027]). We therefore largely refuted hypothesis 5c as well. Only in Kinyarwanda, older students tended to exhibit slightly lower average states. By contrast, we found that the data largely supported hypothesis 5a. Apart from mathematics, school type predicted average motivational states (not reported in table 34). Attending any of the two basic-education schools was associated with lower English competence (ranging from -0.254, [-0.381; -0.116], to -0.498, [-0.609; -0.367]) and cost (ranging from -0.278, [-0.407; -0.139], to -1.426, [-2.061; -0.834]) states and average Kinyarwanda competence states (ranging from -0.193, [-0.358; -0.022], to -0.218, [-0.377; -0.048]). Our results coincided with Bulla et al. (2022b) who also detected that school status shapes school subject-specific motivational beliefs. These differences in motivational levels could be driven by reputational differences between basic-education and boarding schools. As Williams (2019) explains, in Rwanda boarding schools are considered "schools of excellence" (p. 648) whereas basic-education schools are considered to be "for those who failed" (p. 646). These differences in reputation may thus also shape students' general (self-) perceptions. Future research needs to determine of these reputational effects are over and beyond differences in academic achievement (Bulla et al., 2022b).

Limitations and Directions for Future Research

Within the present study, several limitations need to be highlighted. First, the sample was heavily skewed towards students from basic-education schools. As already alluded to, they exhibit some characteristics (e.g., weak cross-subject correlations of school grades) that are at odds with literature (e.g., Möller et al., 2009). A strong focus on other student populations inside and outside Rwanda may yield to different results. Second, within the present study, students reported their motivational states upon completion of lessons (i.e., event-contingent sampling design; Bolger & Laurenceau, 2013). Future replications should therefore test if other time sampling approaches such as signal-contingent designs lead to different result (Finnigan & Vazire, 2018). Third, the current study focused only on competence and cost states excluding other motivational constructs such as intrinsic value (e.g., Eccles & Wigfield, 2020). Future replications should therefore include constructs such as intrinsic value as well. As school subject-specific self-reports, they have partly found to be predicted by prior school grades (e.g., Gaspard et al., 2018). Fourth, school subject-specific and lesson-specific scales were only partly overlapping in terms of their key wording (for example, compare items BE1 and LE3 in table 29). This might have affected the power of tests on convergent validity within the present study. Future tests should therefore employ direct and latent trait scales whose wording of items is overlapping as much as possible.

Some of the present study's implications for future research have already been highlighted above. In addition, future research must also explore some of the implicit properties of average motivational states. First, since they represent traits (Geiser et al., 2013), are average motivational states as stable constructs as theory proposes? This question needs to be studied longitudinally. A second, and potentially more important research question to be addressed in future is about the educational value of average motivational states. Do they predict any educational outcomes such as future school grades over and beyond school subject-specific motivational self-reports (e.g., Finnigan & Vazire, 2018)? To the extent that they are indeed related to important education outcomes, evaluation research may then explore how to best alter average motivational states for the sake of students' educational development in Rwanda and elsewhere.

Conclusion

Unlike school subject-specific competence and cost self-reports, average competence and cost states are only weakly predicted by the GI/E model. This coincides with the latter constructs' low levels of school subject-specificity as well as convergent validity vis-à-vis school subject-specific self-reports. We take this all as evidence that average competence and cost states may be related but distinct constructs from school subject-specific competence and cost self-reports.

5.5. Manuscript four: Does time allocation across classroom activities predict motivational states? Findings from Rwandan secondary schools.

Abstract

Within international development, student-centred practices, including classroom discussions and group work, are considered superior over teacher-dominant practices that focus on rote learning. We used a sample of 868 Rwandan secondary students to study the motivational effects of both activity types as well as teacher explanations. Classroom activities, categorized through external ratings, tended to be teacher centred. Students' motivation was measured through lesson-specific competence, utility, and cost (self-) perceptions repeatedly measured over 13 school days in English, mathematics, and Kinyarwanda, the national language. Latent trait/ state modelling suggested motivational effects to vary across subjects. In mathematics, student-centred activities exerted mainly positive effects. In Kinyarwanda, teacher-dominant activities exerted mainly negative. In English, especially teacher explanations exerted positive effects. The effects of teacher explanations appeared to be mediated by a measure on perceived novelty. Rwandan teachers are therefore best advised to account for students' school subject-specific motivational needs when planning lesson activities.

Keywords: Student-centred pedagogy; verbal interactions; student motivation; classroom observations; latent trait/ state modelling; Rwanda

Introduction

Even though many African countries have made major strides towards universal access to primary and secondary education, they often continue to lag behind in terms of learning outcomes (e.g., Sandefur, 2018). Within the international development discourse on education, student-centred pedagogies have been for long promoted as a way to improve student outcomes (Schweisfurth, 2013; Tabulawa, 2013). Conceptually, student-centred pedagogies are difficult to pin down (Schweisfurth, 2013) but their emphasis on student-student and student-teacher verbal interactions has been considered one of their main pillars (Bremner, 2021). Within the Western hemisphere, classroom practices have been studied not only in terms of their effects on students' learning outcomes but also increasingly in terms of their effects on students' motivation (Eccles, Midgley, et al., 1993; Guthrie et al., 2012; Stipek, 2002; Urdan & Turner, 2005). After all, students' motivation may constitute an indirect route through which classroom practices may influence learning outcomes. Within African contexts, the effects of classroom activities such as student-student and student-teacher interactions have hardly been studied (Bremner et al., 2022; for an exception see Sakata, 2022). This research gap prevents adequately informing educational policymaking on student-centred leaning within Africa. Any study to address this gap should ideally take into account that student motivation often exhibits intrapersonal variability across situations. It implies the notion that motivational experiences vary even within short intervals such as between school lessons. Again, intrapersonal studies on motivation have hardly been carried within the context of Africa (for a rare exception see Bulla et al., 2022c). With the present study, we addressed both research gaps by carrying out an intrapersonal study on the motivational effects of classroom practices using a sample of lower-secondary students from Rwanda. In line with expectancy-value theory, we used a tripartite conceptualisation of student motivation. Thus, we investigated the effects of classroom practices, including student-student and student-teacher verbal interactions, on students' lesson-specific competence, utility, and cost (self-) evaluations repeatedly measured across 13 school days in English, mathematics, and Kinyarwanda, Rwanda's national language.

Operationalizing motivation through expectancy, utility, and cost (self-) evaluations

Within the educational domain, achievement motivation refers to those psychological processes that shape student's learning behaviours (i.e., student engagement) exhibited within academic performance contexts (Wigfield et al., 2006, 2015). Classroom-based engagement in turn has been considered "the direct (and only) pathway" (Skinner & Pitzer, 2012, p. 23) to learning acquisition that is available to students. To the extent that classroom practices shape students' motivational levels, which in turn predict their level of lesson-specific engagement, classroom practices may also indirectly shape students' learning. According to the expectancy-value theory (e.g., Eccles & Wigfield, 2020), students' engagement with a learning task is influenced by three sets of predictors. A first set of predictors concerns students' self-perceptions about how well they are going to perform academically with the upcoming task (i.e., success expectancies). Expectancies in turn are assumed to be influenced through self-evaluations of one's own past academic performance (i.e., "self-concept of ability"; Eccles et al., 1983; Wigfield & Eccles, 2000). Conceptually, both expectancy and self-concept beliefs are two instances of the broader category of competence beliefs. Competence beliefs in turn constitute self-evaluations about one's abilities to meet tasks and situational learning demands (Usher, 2015). Empirically, expectancy and self-concept beliefs were found to be inseparable factors (e.g., Eccles, Wigfield, et al., 1993). Within the present study, we thus use them interchangeably (e.g., Wigfield & Gladstone, 2019) and refer to them as competence self-evaluations. A second set of predictors concerns value perceptions associated with the learning task. Value has been conceptualised multidimensionally (Eccles & Wigfield, 2020). One of its dimensions is utility value, which describes how useful a learning task is perceived to be for working towards one's wider goals inside and outside school (Eccles & Wigfield, 2020). Within education, utility value is particularly of interest because it is the only value dimension that has been found to be malleable even within short time spans (Harackiewicz et al., 2016; e.g., Hulleman, Godes, et al., 2010). A final set of predictors concerns students' perceptions about how cognitively and emotionally taxing engaging with a learning task is (i.e., cost beliefs; Barron & Hulleman, 2015; Flake et al., 2015). Cost beliefs are multidimensional, too. One of its beliefs is effort cost, which refers to the amount of work and attention a task requires until completion (Flake et al., 2015). Within expectancy-value theory, these three motivational constructs are generally considered to be situation specific (Eccles & Wigfield, 2020).This coincides with findings suggesting that the extent to which students experience competence, utility, and costs vary even within short time intervals such as between lessons.

Achievement motivation as intrapersonally variable constructs

Motivational constructs such as competence (e.g., Shavelson et al., 1976) or value beliefs (e.g., Götz et al., 2006) have increasingly been conceptualised hierarchically. These hierarchies are often discussed in terms of traits and states. Whereas traits refer to the stable differences within behaviours, thoughts, and feelings between people, states refer to those concrete behaviours, thoughts, and feelings a person exhibits at a particular time (e.g., Baumert et al., 2017). States are rather short-lived phenomena because they constitute interactions between a person's traits and those contextual factors that make up a specific situation. As soon as these situational factors change, a person may also transition into a new state. Students' motivational levels have been found to vary substantially across situations. A measure of a construct's intrapersonal variability is the intracluster correlation coefficient (ICC). Within a trait/ state mode, the ICC specifies the per centage of variance explained by differences between students (level 2), which in turn implies the per centage of variance explained by differences between situations (level 1; e.g., Geiser, 2020). Intrapersonal variability of motivational constructs within African contexts have not yet been reported in the literature. Within the Western-borne literature, however, ICCs reported were between .27 and .44 in the case of competence self-evaluations (e.g., Ahmed, Werf, et al., 2010; Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021), between .19 and .36 in the case of value beliefs (e.g., Dietrich et al., 2019; Salmela-Aro et al., 2021), and between .28 and .55 in case of cost perceptions (e.g., Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021). Hence, around 45 to 81% of variability within motivational states is due to differences between situational factors rather than person motivational traits. These empirical results thus warrant a trait-state conceptualisation of motivational constructs. They also suggest that motivational constructs such as competence, utility, and even cost perceptions are subjected to situational influences, which may include, at least within the domain of education, classroom practices as well.

The current focus on student-centred pedagogies within the international development discourse

Classroom activities constitute those "opportunities to learn" (Jansen, Scherer, et al., 2015, p. 15) students engage with during a lesson. Within the international-development discourse on education, there has been for long an emphasis on the student-centred organization of classroom organization over teacher-centred classroom practices (e.g., Bremner et al., 2022; Schweisfurth, 2013). For example, the launch of Education for All (EFA) in 1990 embodied an agreement by international-donor agencies that the most effective pedagogies would be “learner-centred” (Haddad et al., 1990, p. 68). Ever since, educational reforms in Africa (e.g., Frost & Little, 2014; O’Sullivan, 2004; Sakata, 2022; van de Kuilen et al., 2019) and elsewhere (e.g., Bremner et al., 2022) have stressed student-centred pedagogies. Within student-centred approaches, students are the main actors within learning processes who take responsibility for their own learning progress. Accordingly, teachers assume a predominantly faciliatory role within classrooms in the sense that they help to adjust learning processes to students’ prior knowledge and current needs (Bremner, 2021). By contrast, teacher-centred approaches are organized around the teacher who controls the pace and content of the learning process in form of whole-class instructions often derived from textbooks. Within these whole-class processes, students are mainly passive recipients of knowledge imparted by the teacher (Smit et al., 2014). Arguments put forward in support of student-centred approaches implied that they help students to learn more effectively, to develop critical thinking as well as creativity, and to practice principles that have been otherwise associated with democratic citizenship (Schweisfurth, 2013). A first review indeed suggests generally positive cognitive as well as non-cognitive outcomes associated with student-centred pedagogies within low- and middle-income countries (Bremner et al., 2022). However, the authors also point out that the existing evidence is “provisional” at best and that more research on the effects of student-centred classroom practices within low-resource countries is warranted. Current research on student-centred pedagogies may also be hampered by the fact that they constitute complex concepts (Bremner, 2021; Schweisfurth, 2013). Studying their various aspects in isolation may help to identify what elements of student-centred pedagogies are most relevant to student outcomes. One of most common principles associated with student-centred pedagogies has been called “interaction” (Bremner, 2021, p. 166), “dialogue” (Schweisfurth, 2013, p. 146) or “active teaching” (Ngware et al., 2012) and refers to the provision of opportunities to students to verbally interact with each other or with the teacher during lessons (Bremner, 2021). As the next section highlights, they have also been found to predict motivational variables.

Classroom discussions and group work as predictors of lesson-specific motivation

Verbal interactions within classrooms appear also in form of whole-class discussions and group work. The former can be loosely defined as “all verbal exchanges where one individual addresses another individual or individuals and at least one addressed individual replies” (Howe & Abedin, 2013, p. 325). It is implied that through classroom discussions students feel they have a say in learning activities and thus can contribute to their own learning process. Discussions thus can be a strategy to improve students’ autonomy (Deci & Ryan, 2000), a state which has been associated with feelings of competence and interest. The former resonates with Jansen et al. (2015), wo found classroom discussions to positively predict students’ competence experiences within science lessons. It has been argued that especially situational interest positively influences students’ attentional and retrieval processes (Hidi, 1990), which in turn should reduce cost perceptions around learning activities. Small-group work implies students being divided into groups where they work together to complete a given task through verbal interactions (e.g., Murphy et al., 2009). Compared to whole-class discussions, perceived autonomy may thus even be lager within small-group settings, with positive effects on students’ interest levels (X. Wu et al., 2013) and thus cost perceptions. In small-group settings, individuals that work together can pool their mental resources and abilities, which should lead to higher confidence levels of being able to successfully complete the task (i.e., group efficacy; Bandura, 2000) compared to settings where students work individually. Also, especially when performing challenging tasks, working in groups requires less mental effort from its members in order to attain given learning outcomes (i.e., "efficiency"; F. Kirschner et al., 2009, p. 615). Hence, incorporating group work into lessons may also lead to more favourable competence and cost (self-)evaluations amongst students. Both classroom discussions and group work may also have positive effects on utility perceptions around lesson content. As already pointed out, utility perceptions are especially shaped through significant others such as teachers and peers (Harackiewicz et al., 2012b; Wigfield et al., 2020). Classroom interactions may thus provide students with the opportunity to enquire about the instrumental value of lesson content. It might have been for this reason that Urdan and Turner (2005) make the case for "active participation" (p. 303) in classrooms as a way to promote the perceived task value of learning. These predictions have hardly been tested within the African context. In Tanzania, Sakata (2022) carried out classroom observations during 17 lessons collected data on of student attitudes towards learning including confidence and motivation, the latter being a proxy for value perceptions. She did not find classroom activities associated with student-centred pedagogies to predict any of those (self-) evaluations. However, she did not take into account the intrapersonal variability within student motivation, something done within the present study,

Teacher-centred activities as predictors of lesson-specific motivation

Despite a policy impetus towards student-centred learning, teacher-centred practices have tended to persist in many low- and middle-income countries (Nag, 2014; Schweisfurth, 2013). Using classroom observations, Sakata (2022) found teacher-centred activities to account for 81% of the lesson time at Tanzanian primary schools whereas student-centred activities accounted only for about 14%. Frost and Little (2014) found teacher-oriented activities accounted for around 75% and student-centred activities for another 15% of time observed at Ethiopian primary schools. Especially within low- and middle-income countries, teacher-centred pedagogies are not only widespread but also often emphasizing surface (rote) learning (Nag, 2014). Within classrooms, this often manifests itself in form of students copying large amounts of notes from the blackboard and then to practice what they copied through a series of individual seatwork exercises (e.g., Honeyman, 2015b). In Ethiopia, for example, Frost and Little (2014) found copying notes and individual seatwork to jointly account for around 28% of classroom time whereas classroom discussions and group work only accounted for around 4%. Within the present study, we consider both note copying and individual seatwork as indicators for teacher centred classroom practices in its most "dominant" (Ngware et al., 2012) form. Especially prolonged periods of copying notes and individual seatwork exercises, indiscriminately administered to all students in class, may be perceived as boring and potentially meaningless by students ("autonomy-suppressing"; Assor et al., 2002). To the extent that perceived autonomy can be associated with feelings of competence and favourable cost evaluations, suppressed autonomy coming along with note copying and individual seat work may thus lead to the feeling of lack of competence and unfavourable cost evaluations ("a-motivation"; Assor et al., 2005). Both teacher-dominant classroom activities may also imply stifling students' active participation in classrooms. To the extent that active participation is indeed a way to promote task value amongst students (Urdan & Turner, 2005), copying notes and individual seatwork should thus also entail negative effects on perceived daily utility of lesson content. Teacher-centred approaches towards learning are not associated only with negative motivational effects. Especially within the direct-instruction paradigm (P. A. Kirschner et al., 2006), the importance of teacher explanations has been emphasized. Whereas classroom discussions and group work are bi- or multidirectional interactions, teacher explanations are “one-way teacher-to-learner transmission” (Schweisfurth, 2013, p. 147). Here, teachers use clear guided explanations that carefully walk students through new content and help them to internalize the key learning points (Rupley et al., 2009; Winograd & Chou, 1988). In this respect, teacher explanations provide students with a structure that is “competence supportive” (Vansteenkiste et al., 2012, p. 432). Ideally, teacher explanations provide students with guidance and directions that are aligned with their needs and problems during the learning process. This in turn may lead to more favourable cost and potentially daily-utility perceptions of learning content. Hence, teacher-centred pedagogies in form of copying notes and individual seatwork may have negative motivational effects whereas teacher-centred activities in form of teacher explanations may have positive effects on students’ competence, utility, and cost perceptions. Within the present study, we investigated these effects using a sample from Rwanda, Sub-Saharan Africa.

The Rwandan School Context

Rwanda is a low-income country in East Africa with Kinyarwanda as its national language (NISR, 2016). In 2012, 12 years of free basic education (12YBE) was introduced. Students now first attend six years of primary schooling before going for three years of lower (S1 – S3) and then upper secondary education (S4 – S6). Traditionally, boarding schools have been providing secondary education to Rwandans. They still have the reputation of being “schools of excellence” (Williams, 2019, p. 648), which may be explained through the facts that they continue to attract best performing students and that they charge school feels, which allows them to provide better learning and teaching infrastructure. Since those fees are not affordable by many Rwandan households, the government introduced basic-education schools to absorb the increasing numbers of secondary students due to 12YBE. In 2016, Rwanda also adopted student-centred pedagogy as the official pedagogy. Accordingly, Rwanda's Ministry of Education (MINEDUC, 2015) has also been promoting student-centred approaches, which include "participatory and interactive methods that engage young people in the learning process" (p. 63) and accordingly expected teachers to be mainly facilitators of students' learning processes (Rwanda Education Board 2015a, 2015b). Even though "chalk and talk" (i.e., rote learning) continues to be a common teaching method in Rwandan classrooms (e.g., DeStefano et al., 2012), anecdotal evidence suggests learner-centred pedagogies are not entirely uncommon either. van de Kuilen et al. (2020) carried out classroom observations amongst others of 12 secondary teachers. In all classes observed, student-teacher and student-student interactions were promoted. As they report, this manifested especially through group work that teachers initiated. They also observed classroom discussions in the form of "whole class questioning during which the teacher, with the help of hints and clues, enabled learners to discover new knowledge themselves" (p.11). However, their sample was small and limited towards boarding schools and skewed towards teachers that were "good implementor" of student-centred pedagogies. More research on the extent of the different types of teaching practices within Rwandan classrooms is therefore warranted. Given the government's impetus on student-centred pedagogies, the effects of classroom practices on students' motivation need to be examined as well. Any such examination should also control for age and gender. According to Bulla et al. (2022b), older students tended to have lower competence beliefs and more negative cost beliefs while female students tended to have lower competence beliefs, higher utility perceptions, and more negative cost beliefs.

The present study

In the present study, we examined the effects of classroom practices associated with student- and teacher-centred pedagogies on Rwandan students' competence, utility, and cost (self-) perceptions. Classroom activities can be distinguished not only in terms of the specific tasks (e.g., group work versus individual seatwork) they imply but also in terms of the quality by which each of these tasks are implemented in class (e.g., Bremner, 2021). Within the present study, we adopted an admittedly simple focus by focussing on the allocation of lesson time across different classroom activities and the implications it may have for students' lesson-specific motivation. The study was guided by the following research questions and hypotheses.

1) *To what extent are the student-centred and teacher-centred lesson activities used within Rwandan classrooms?* In line with the studies reported by Sakata (2022) and Frost and Little (2014), we expected teacher-centred activities to account for the biggest part of lesson time observed at Rwandan schools. Reflecting especially the anecdotal evidence provided by van de Kuilen et al. (2020), however, we expected the share of lesson time to fall on teacher-centred activities to be less than those 75 to 81% that Sakata (2022) and Frost and Little (2014) reported (hypothesis 1).
2) *To what extent does the time allocated across student-centred and teacher-centred classroom activities predict Rwandan students' lesson-specific expectancy, utility value, and cost (self-) evaluations?* Within the present study, we operationalised student-centred focus through the allocation of time to classroom discussions and group work. We hypothesized that their use positively predicts students' lesson-specific expectancy, utility, and cost (self-) perceptions (hypothesis2a). By contrast, classroom activities in form of note taking and individual seatwork were taken as instances of a dominant teaching style

(Ngware et al., 2012) and assumed to negatively predict expectancy, utility, and cost (self-) perceptions (hypothesis2b). Finally, even though teacher explanations are teacher-centred (but not teacher-dominant) practices, we assumed the time allocated to teacher instructions would exert positive effects on all three motivational states (hypothesis2c).

3) *To what extent are Rwandan students' experiences of expectancy, daily utility, and cost predicted by age and gender?* In line with Bulla et al. (2022b), we expected that both age and gender partly influence expectancy, daily utility, and cost (self-) evaluations of students. To be more precise, we expected older students to have lower competence beliefs and more negative cost beliefs and female students to have lower competence beliefs, higher utility perceptions, and more negative cost beliefs (hypothesis 3; also see Bulla et al., 2022b).

Method

Procedures

Within the present study, we used data that was collected as part of a wider project on student motivation in Rwanda (Bulla, Dominik et al., 2022; e.g., Bulla et al., 2022b, 2022c). The data on lesson-specific competence, daily-utility, and cost states were collected intraindividually for 13 consecutive days in February and March 2020 targeting the subjects of English, and mathematics using a sample of lower secondary students (S1 – S3) enrolled at three public schools inside and outside Kigali. Two out of those three schools were basic-education schools (A2 and B1). The last one (B2) was a boarding school. Across those 13 days and three schools, student ratings and classroom observations were collected for 313 lessons in total. We removed those lessons where we observed students to only copy notes

(12 lessons) or to write exams (7 lessons). Within each school subject-specific dataset, we removed those students for whom we had not collected at least three ratings for the given subject. This resulted in the exclusion of another 20 lessons across the three subjects. We ended up using 274 of the initial lessons covered (101 in English, 49 in Kinyarwanda, and 124 in mathematics). As table 35 highlights, we did not manage to include any Kinyarwanda lessons at boarding school B2. Table 35 also indicates the average length of a lesson in any of the three subjects expressed in minutes.[27] By and large, Kinyarwanda lessons appeared to be slightly shorter. Not to disrupt school lessons, self-reports were collected in the last five minutes of lessons (i.e., event-contigent design; e.g., Bolger & Laurenceau, 2013). The self-report questionnaire was translated into Kinyarwanda and also elicited personal information (i.e., name, age, gender, class), which was then used to track student responses across lessons. Across the 274 lessons, 8652 lesson-specific self-reports were gathered (3242 in English, 1411 in Kinyarwanda, and 3999 in mathematics). Each school subject-specific dataset was imbalanced since students varied in terms of the average number of self-reports provided. It was 4.62 ($SD = 1.36$) in English, 3.77 ($SD = 0.75$) in Kinyarwanda, and 5.17 ($SD = 1.76$) in mathematics. All questionnaires were entered independently by two separate data-entry clerks. Mismatches during data entry between both clerks were resolved by reverting back to the hard copies of the questionnaires. For students to partake in the study, they were asked to sign a child assent form. Head-teacher consent was sought as they acted *in loco parentis* (British Psychological Society, 1992). As a token of appreciation, all students that submitted consent forms participated in a tombola carried out after field work. Per class, three students won a pocket calculator worth around 75p. The study has ethics approval from University of Oxford (ED-CIA-19-084) and University of Rwanda (266/CMHS IRB/2018).

[27] In Rwanda, a lesson period is supposed to take 40 minutes. In case of double or triple lessons, however, the length of a lesson can be longer than 40 minutes. Also, teachers did not always start or ended lessons on time, which added additional variability to average lesson lengths.

Table 35: *Distribution of lessons attended during data collection (manuscript four)*

School/lesson-specific statistics	English	Kinyarwanda	Mathematics	Total
A2	21	12	42	75
B1	57	37	52	146
B2	23	0	30	53
Total	101	49	124	274
Minutes (SD)	51.6 (21.8)	43.8 (21.7)	55.7 (23.8)	51.8 (23.0)
Average ratings/ lesson (SD)	81.4 (37.8)	60.1 (26.1)	89.8 (39.9)	81.4 (38.4)

Note. 'Minutes' refers to the length of lessons expressed in minutes. 'SD' refers to standard deviations.

Participants

In total, we used a sample of 868 students (702 in English; 374 for Kinyarwanda; 773 in mathematics). Only 222 provided at least three self-reports within each of the three subjects. The full sample was distributed across 22 classrooms between the three schools (see table 36). At school A2, we did not carry out classroom observations within S1. With almost 100 students within each class, we considered them too big for classroom observations. Despite the removals of the two classes from the study, school A2 continued to have the highest average class size. The boarding school had the lowest average number of students in class. The average number of study participants per class was 39.5 ($SD = 8.58$) and ranged from 25 to 56 students. Table 36 provides additional information on sample characteristics. At boarding school B2, fewer females tended to participate than at the basic-education schools. Boarding-school students also tended to be younger than basic-education students.

Table 36: *Descriptive school statistics (manuscript four)*

School characteristics	A2	B1	B2
Boarding school?	No	No	Yes
Number of classes	7 (5)	11	6
Average class size (SD)	60.6 (9.11)	48.55 (7.02)	43.2 (8.55)
N (% female)	238 (55.46)	416 (56.73)	214 (42.06)
Age (SD)	15.47 (1.59)	15.32 (1.65)	14.13 (1.36)

Note. 'N' refers to sample size. 'SD' refers to standard deviations. At school A2, we only used data from 5 of the 7 classes that existed.

Measures

Lesson-specific self-reports

To measure lesson-specific (self-) perceptions, we employed scales that were developed for the Rwandan school context. To do so, we largely followed the scale-development process outlined in authors (2022b). Lesson-specific competence, daily utility, and cost perceptions were each measured through three items (see table 37). Students responded to each item using a three-options response format consisting of 1) yes, very much; 2) A bit; or 3) no, not at all. All items were coded in such a way that higher values represent more favourable perceptions. Table 38 presents average scores for each scale, subject, and scale. Compared to basic-education students, boarding-school students appear to have slightly higher average scores in case of competence and cost (self-) perceptions. In terms of daily utility, they are partly lower. For each of the three school subjects, table 39 presents within-level polychoric correlations for lesson-specific scales and single-level polychoric correlations for domain-specific motivational scales as measures of instrument reliabilities. Correlations were substantial at the within-person level but particularly high at the between-person level. Together, we took this as evidence of acceptable scale reliabilities. We also included a question on whether or not students felt they learned anything new during the lesson, which had the same answer format as the self-report items. As table 38 shows, novelty was reported most within mathematics and least in Kinyarwanda. We also noted that perceived novelty was higher amongst boarding-school (μ_{NBO}= 2.54; SD_{NBO} = .74) than basic-education students (μ_{NBA}= 2.44; SD_{NBA} = .79).

Table 37: The lesson-specific motivational scales (manuscript four)

	Item
	Competence scale
LC1	Do you think you understood today's activities of [SUBJECT]?
LC2	Do you think you were capable to do today's activities of [SUBJECT]?
LC3	Do you think you understood today's activities of [SUBJECT]?
	Daily-utility scale
DU1	Do you think you will use what you learned in today's activities in [subject] in your everyday life?
DU2	Do you think today's activities of [SUBJECT] will help you in your daily life?
DU3	Do you think today's activities of [subject] will be useful for your everyday life?
	Cost scale
LE1	Were today's activities of [SUBJECT] hard?
LE2	Were today's activities of [SUBJECT] tiring?
LE3	Were today's activities of [SUBJECT] confusing?

Classroom activities

Through systematic classroom observations (Salvia et al., 2007; Sanders et al., 2010; Wragg, 2012) by six external raters located in class, we measured the amount of time allocated to a given classroom activity (Griskell et al., 2020). To allow teachers and students to settle in, in-class observations began two minutes after a lesson had started. Observations followed a round-robin design (Ocumpaugh et al., 2015) that involved observing classroom activities for five second every 30 seconds. As table 35 highlights, we collected 81.4 ratings on average across all lessons with slightly more ratings in mathematics than in English and Kinyarwanda. The observation routines in class were exactly timed by means of a multi-timer app installed on a smart phone each rater was equipped with. Through audible signals communicated via headphones, raters knew when to start and stop an observation period. Each of the five seconds were coded based on the classroom activity that accounted for the biggest part of those five seconds (i.e., predominant-activity sampling; Hutt & Hutt, 1970, p. 69). Classroom activities were categorized based the World Bank's (2015) revised version of Stallings' (e.g., 1977) Classroom Snapshot Observation System. We adjusted it to the

Rwandan context by slightly reworking the activity categories during an extensive pilot phase. As a result, we ended up with 11 categories grouped into instructional, managerial, and miscellaneous activities. We removed five of them because on average they just accounted for not more than 1.9% of the time across the 274 lessons we focused on. For example, classroom management accounted for only 5.3% (standard errors = 6.6) of the time across all lessons. In addition, we collapsed two categories into one ("individual seatwork") because the two classroom activities referred to by these two categories appeared to be rather similar in terms of their behavioural implications for students. Table 40 lists the remaining five classroom activities we focussed on within the present analysis.

Table 38: *Average score per scale (manuscript four)*

Scale	A2 Basic	B1 Basic	B2 Boarding	All
		English		
Competence	2.34 (0.53)	2.57 (0.49)	2.72 (0.46)	2.56 (0.51)
Daily utility	5.14 (1.5)	4.48 (1.67)	4.16 (1.35)	4.54 (1.6)
Cost	2.23 (0.52)	2.35 (0.57)	2.52 (0.55)	2.36 (0.57)
'New'	2.57 (0.71)	2.40 (0.82)	2.44 (0.78)	2.45 (0.79)
		Kinyarwanda		
Competence	2.52 (0.47)	2.77 (0.4)	NA	2.7 (0.43)
Daily utility	4.24 (1.67)	5.13 (1.68)	NA	4.9 (1.72)
Cost	2.28 (0.56)	2.5 (0.59)	NA	2.44 (0.59)
'New'	2.39 (0.82)	2.37 (0.84)	NA	2.38 (0.84)
		Mathematics		
Competence	2.39 (0.52)	2.42 (0.54)	2.55 (0.48)	2.44 (0.52)
Daily utility	3.94 (1.65)	4.5 (1.76)	4.29 (1.65)	4.25 (1.71)
Cost	2.24 (0.53)	2.2 (0.57)	2.26 (0.55)	2.23 (0.55)
'New'	2.46 (0.77)	2.47 (0.78)	2.61 (0.69)	2.50 (0.76)

Note. 'Basic' = basic-education school; 'Boarding' = boarding school. Figures in brackets represent standard deviations.

All raters employed completed a 15-day long training prior to the study. Throughout study implementation, they were all subject to regular reliability checks. During those checks, the first author accompanied a rater unannounced and then in class simultaneously but independently performed the same rating as the rater. This allowed determining average interrater reliability using Cohen's Kappa (1960). The average agreement rate was 0.80 [.77; .82], indicating substantial agreement between raters (Landis & Koch, 1977). For each lesson, we aggregated the individual ratings to determine the approximate per centage of lesson time allocated to each of the different classroom activities.

Statistical Analysis

Latent Single Trait Multiple States (STMS) Modelling

To account for the imbalanced and hierarchical (i.e., lesson-specific self-reports nested with students) nature of the data (Bolger & Laurenceau, 2013), we performed multilevel structural equation modelling in order to investigate relations between time allocated across the different classroom activities and subsequent students' lesson-specific motivational self-reports. To ensure convergence of models, we performed the analysis separately on each of the three school subject-specific data sets. Within each school subject-specific setup, each of the three motivational factors was specified through a multiple-indicator latent single-trait multiple state (STMS) model (Castro-Alvarez et al., 2021; Geiser et al., 2013 also figure 16). In this setup, lesson-specific self-reports at level one can be interpreted as time-point specific deviations (i.e., states) from person-specific averages (i.e., the traits) located at level two. State residuals as well as error variables were implied to have zero means and to be uncorrelated across time. We did not consider our average number of lesson-specific self-reports large enough for any autocorrelational analysis (Schultzberg & Muthén, 2018). State residuals were specified to be uncorrelated with latent traits. Model intercepts and factor loadings were all assumed to be time invariant (i.e., "pure state-variability model"; Geiser et al., 2013, p. 5). In line with the assumption of cross-level measurement invariance (Stapleton et al., 2016), factor actor loadings were also constrained to be equal between levels. For both states and traits, the first factor loading was fixed to one. To study the effects of classroom activities (see the β-coefficients in figure 16), we included five continuous variables as level-two predictors, which represented the per centage of teaching time allocated across the five activity types (jointly, they did not sum up to 100%). Within these latent regressions, we also added a set of additional control variables at level one (not included in figure 16). One was about whether or not students felt they learned something new. We also included a set of dummies representing weekdays ('Monday' as reference), two dummies representing the lesson periods[28] with the early-morning block as the reference. School dummies (boarding school as reference) as well as age and gender were included as predictors of the student-specific motivational traits at level two. Hence, age and gender effects were modelled as level-two coefficients.

Table 39: *Average within-scale polychoric correlations (manuscript four)*

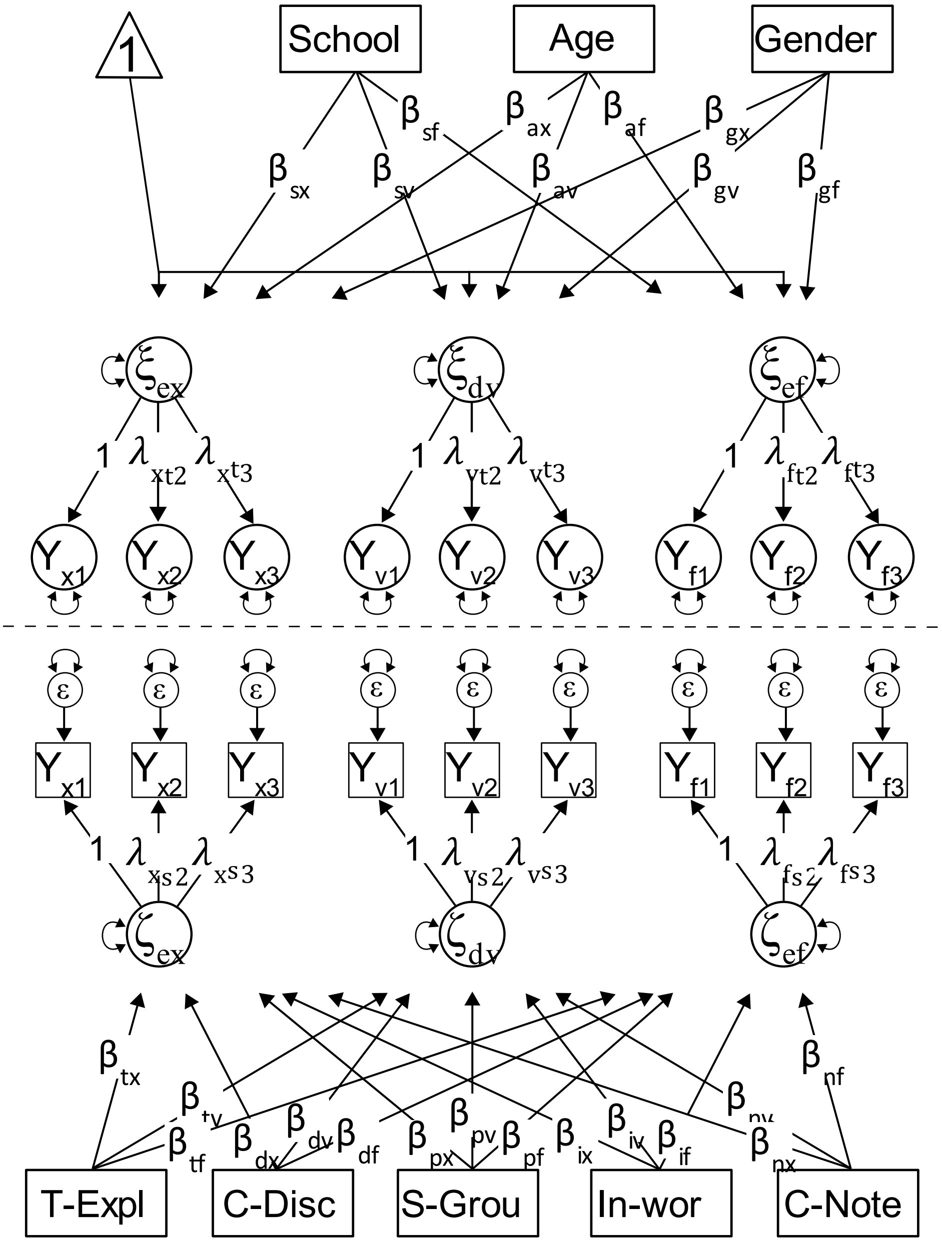
Scale	English	Kinyarwanda	Mathematics	Average
average within-level polychoric correlations (lesson-specific self-reports)				
Competence	.59	.54	.62	.58
Daily utility	.52	.55	.58	.55
Cost	.52	.54	.52	.53
Average	.54	.54	.57	
average between-level polychoric correlations (person-specific self-reports)				
Competence	.98	.96	.97	.97
Daily utility	.96	.98	.98	.97
Cost	.97	.98	.95	.97
Average	.97	.97	.97	

Note. Polychoric correlations were estimated using a multilevel framework. Level one represented the lesson-specific within-person level and was nested within level two which constituted the between-person level. The correlational analyses concerning the lesson-specific measures were performed using the 'TYPE = TWOLEVEL BASIC' command. The correlational analyses concerning the school subject-specific direct trait measures were performed using the 'TYPE = COMPLEX BASIC' command.

[28] By and large, a typical Rwandan school day consists of 9 lesson periods, divided into three blocks with three periods each. There is a 20-minute break between (early-morning) block 1 and (late-morning) block 2 and a 60-minute lunch break between block 2 and the afternoon block 3. This structure may vary across schools at times.

*Figure 16**:** The hypothesized structural model (manuscript four)*

The between-student level



The lesson-specific level

Note. It is adapted from Geiser et al. (2013) and used to estimate level-1 structural relations between classroom activities and motivational states. ‘T-Expl’ refers to teacher explanations; ‘C-Disc’ refers to whole-class discussions; ‘S-Grou’ refers to small group work; ‘In-wor’ refers to individual seat work; and ‘C-Note’ refers to copying notes. At the lesson-specific level (level 1), we also controlled for day of the work (four dummies with Monday as reference), time of the school day (two dummies representing the late-morning lesson block and the afternoon lesson block with the early morning block as reference) and novelty. To avoid cluttering, they are not included here. At the between-student level, school was operationalised through two dummies with boarding school as reference. Motivational states as well as motivational traits were allowed to be correlated with each other but not across levels.

Table 40: *The classroom-activity coding scheme (manuscript four)*

Activity	Definition
Teacher explanations	The teacher is explaining or demonstrating to the class tasks, homework, or any other academic content with the intention to make students understand it.
Group work	All students in the class are divided into groups where they are expected to work on a group assignment. In most cases, students rearrange their seats in accordance with how they have been divided into groups. During group work, the teacher generally goes from group to group to provide support and guidance.
Discussions	A student is verbally engaging with the teacher or other students to ask questions, to answer questions, or to discuss academic content while the rest of the class is listening.
Copying notes	Students are copying notes that the teacher is writing on the blackboard. At times, the teacher may read out passages written on the board mainly to help students copy notes properly.
Individual work	All students in class are working on a given task by themselves. In some cases, a student is at the blackboard to work the task there, potentially for demonstration purposes.

Note. Coding Scheme adapted from World Bank (2015). Also see Stallings (1977)

Bayesian Estimation

We run all models using the two-level Bayesian estimation framework in Mplus version 8.7 (L. K. Muthén & Muthén, 2017). Seed value chosen was 21012020. The posterior distribution was obtained using the iterative Markov chain Monte Carlo (MCMC) algorithm (van de Schoot et al., 2014) using two chains and two processors used for parallel computing. As for the potential scale reduction (PSR) factor, we specified the MCMC convergence criterion to be 0.01 (Hox et al., 2012). We discarded the first 5000 iterations as 'burn-in'. To reduce the effects of parameter autocorrelation between subsequent chain iterations, we only used every 30th iteration (thinning) during the computation of the final posterior distribution. We rerun every model with the number of iterations twice the number of iterations required during the first run. During all model estimations, we used Mplus' diffuse priors (for their hyperparameters see L. K. Muthén & Muthén, 2017). Given a sample size between 374

(Kinyarwanda) and 773 (mathematics), we expected the use of the default priors as less consequential. To perform convergence diagnostics, we inspected the trace plots. We checked whether or not all chains converged to the same target distribution and whether or not all iterations were based on stable chains. Model fit itself was assessed through the posterior predictive p-value (PPP), which are computed using the likelihood-ratio chi-square statistic (Gelman, 2014). Posterior predictive p-values (PPP) of 0.5 as well as a confidence interval for the chi-square statistic that contains the zero value are taken as evidence of excellent model fit. Model fits generally turned out to be acceptable (PPP-values between .143 and .436, see table 42 and 44) with all 95% confidence intervals containing the zero value.

Table 41: Factor means, variances and ICCs (manuscript four)

Scale	Within-level variance	Between-level mean	Between-level variance	ICC
		English		
Comp	**2.31 [1.81; 2.94]**	**2.76 [.97; 4.43]**	**3.86 [2.96; 5.03]**	.63
Daily	**1.25 [.96; 1.62]**	**3.04 [1.38; 4.78]**	**2.96 [2.25; 3.90]**	.70
Cost	**1.53 [1.25; 1.90]**	1.12 [-.86; 4.56]	**2.63 [2.11; 3.31]**	.63
		Kinyarwanda		
Comp	**1.96 [1.33; 3.11]**	**2.76 [.82; 5.55]**	**3.34 [2.20; 5.35]**	.63
Daily	**1.60 [1.12; 2.40]**	**2.75 [.85; 4.69]**	**3.94 [2.70; 5.96]**	.71
Cost	**2.25 [1.57; 3.39]**	1.95 [-.15; 3.98]	**4.50 [3.11; 6.75]**	.67
		Mathematics		
Comp	**2.40 [1.98; 2.93]**	**2.81 [.97; 4.91]**	**2.77 [2.22; 3.47]**	.54
Daily	**1.53 [1.27; 1.85]**	**2.94 [.83; 4.20]**	**4.23 [3.41; 5.27]**	.73
Cost	**1.51 [1.27; 1.82]**	.80 [-.97; 2.48]	**1.73 [1.42; 2.13]**	.53

Note. ‘Comp’ = competence states; ‘Daily’ = daily-utility states; ‘Cost’= cost states. ‘ICC’ = intracluster correlation. Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits: English: Sample size = 702; PPP = 0.422, f (95% CI) = -36.42 - 44.42; Kinyarwanda: Sample size = 374; PPP = 0.436, f (95% CI) = -35.76 - 45.26; mathematics: Sample size = 773; PPP = 0.180, f (95% CI) = -24.84 - 66.49. Coefficients in bold are significantly different from zero.

Results and discussion

Table 41 on the previous page presents between-level means, the within- as well as between-level variances, and the resulting ICCs for each of the motivational constructs. ICCs were rather high with an average of .60 for competence, .71 for utility, and .61 for cost (self-) perceptions. This implies that around 29 to 40% of the variance within motivational factors originated at the lesson-specific level. Table 42 presents the factor correlations. Between-person correlations were consistently higher than within-person correlations. Given the substantial intrapersonal variability and the correlational patterns that varied across levels, we concluded that a trait/state operationalization of the three motivational constructs was indeed warranted.

Table 42: Standardised factor correlations (manuscript four)

Scale	Comp	Daily	Cost
		English	
Comp	-	0.58 [.50; .65]	0.76 [.71; .80]
Daily	0.54 [.47; .61]	-	0.46 [.37; .54]
Cost	0.58 [.53; .64]	0.24 [.10; .37]	-
		Kinyarwanda	
Comp	-	0.66 [.56; .75]	0.78 [.70; .84]
Daily	0.58 [.47; .68]	-	0.48 [.36; .59]
Cost	0.50 [.40; .60]	0.16 [.04; .38]	-
		Mathematics	
Comp	-	0.52 [.45; .59]	0.69 [.64; .75]
Daily	0.45 [.39; .50]	-	0.39 [.30; .47]
Cost	0.61 [.57; .65]	0.22 [.15; .28]	-

Note. ‘Comp’ = competence states; ‘Daily’ = daily-utility states; ‘Cost’= cost states. Coefficients below the diagonal represent within-person (level 1) correlations. Coefficients above the diagonal represent between-person (level 2) correlations. As for model fits, see table 42. All coefficients are significantly different from zero.

To what extent are the student-centred and teacher-centred lesson activities used within Rwandan classrooms?

Table 43 presents descriptive statistics on the average distribution of classroom activities across schools and subjects. On average, our final coding scheme categorized 90.38% (SD = 12.96) of lesson time observed. The classroom activities associated with dominant teaching styles accounted for the largest share of lesson time (42%), followed by teacher explanations (27%), and student-centred activities (22%). Despite the policy impetus towards student-centred learning (MINEDUC, 2015), teacher-dominant practices thus continue to be most common teaching method within Rwandan classrooms. Given Rwanda's cultural environment, this finding may not surprise. Honeyman (2015b) observed "obedience, hierarchy, and respect for authority" (p. 24) to be an integral part of Rwandan culture. Within the educational domain, the teacher has traditionally been considered "the central arbiter of knowledge" (p. 161) and therefore classroom learning used to be organized around "teachers teach; students listen" (p. 136), which had manifested itself in form of "chalk and talk" (e.g., DeStefano et al., 2012) at least in the past. Nonetheless, our results also suggests that within Rwandan classrooms it is not only the teacher anymore who just talks while students just listen. In fact, compared to Ethiopian (Frost & Little, 2014) and Tanzanian (Sakata, 2022) primary schools, Rwandan secondary schools seem to be less teacher centred and more student oriented. We therefore considered hypothesis 1 as supported by the data. Only in about 7.97% of lessons we did not find any activities we classified as student oriented, which implies that within the majority of lessons at least some teaching time was allocated to classroom discussions and group work. This coincided with findings by van de Kuilen et al. (2020) that implied dialogic classroom activities are generally implemented across classrooms at the secondary level. Unlike van de Kuilen et al. (2020), however, we found classroom discussions the more dominating student-oriented classroom activity than group work. According to correlational analyses (not reported here), the time allocated to the different classroom activities tended to be weakly but negatively correlated. This suggests that teachers generally did not combine certain classroom activities. However, we found a moderate and positive correlation between the amount of time allocated to teacher explanations and students' perception of having learning something new during class (novelty) in English (ρ = .43; p < .000) and mathematics (ρ = .27; p < .000). This suggests that new learning topics are introduced to students especially through teacher explanations.

Table 43: *The extent of classroom activities observed (manuscript four)*

School		English	Kinyarwanda	Mathematics	All
			Teacher explanations		
A2	BA	48 (25)	23 (14)	23 (13)	30 (20)
B1	BA	18 (12)	31 (22)	31 (15)	26 (17)
B2	BO	21 (11)	NA	26 (17)	24 (15)
all		25 (19)	29 (20)	27 (15)	27 (18)
% without		4.67	4.76	4.58	4.65
			Classroom discussions		
A2	BA	17 (15)	21 (20)	14 (11)	16 (14)
B1	BA	14 (16)	18 (15)	11 (09)	14 (14)
B2	BO	22 (19)	NA	08 (09)	14 (16)
all		17 (17)	18 (16)	12 (10)	15 (14)
% without		16.82	6.35	9.16	11.30
			Small group work		
A2	BA	00 (01)	05 (10)	02 (05)	02 (06)
B1	BA	12 (23)	09 (18)	09 (17)	10 (20)
B2	BO	05 (08)	NA	10 (22)	08 (17)
all		08 (18)	08 (16)	07 (16)	07 (17)
% without		74.77	65.08	72.52	71.76
			Note copying		
A2	BA	12 (22)	14 (16)	10 (12)	11 (16)
B1	BA	17 (24)	21 (32)	13 (14)	17 (24)
B2	BO	16 (18)	NA	12 (25)	14 (22)
all		16 (22)	19 (29)	12 (17)	15 (22)
% without		39.25	41.27	32.82	36.88
			Individual seatwork		
A2	BA	15 (20)	29 (29)	45 (20)	34 (25)
B1	BA	21 (26)	11 (14)	27 (24)	20 (24)
B2	BO	28 (18)	NA	39 (28)	34 (25)
all		21 (24)	16 (20)	36 (25)	27 (25)
% without		26.17	30.16	8.40	19.27

Note. 'BA' = basic-education school; 'BO' = boarding school. '% without' refers to the share of lessons where the corresponding activity was not observed at all. Numbers in brackets represent standard errors.

Table 44: The estimated coefficients of the structural models (manuscript four)

Classroom activity	English Competence	English Daily utility	English Cost	Kinyarwanda Competence	Kinyarwanda Daily utility	Kinyarwanda Cost	Mathematics Competence	Mathematics Daily utility	Mathematics Cost
Within students (level one)									
TEX	**0.19**	**0.11**	**0.19**	**-0.26**	-0.10	-0.05	0.42	0.04	**0.10**
	[.10; .28]	**[.01; .22]**	**[.10; .28]**	**[-.43; -.08]**	[-.29; .10]	[-.21; .12]	[-.05; .13]	[-.07; .15]	**[.01; .19]**
GWO	0.02	0.00	0.03	0.05	-0.06	0.15	**0.10**	0.06	**0.11**
	[-.06; .10]	[-.09; .10]	[-.05; .10]	[-.12; .22]	[-.25; .13]	[-.01; .31]	**[.01; .19]**	[-.05; .17]	**[.02; .21]**
DIS	0.05	**0.12**	**0.14**	0.01	**-0.21**	**0.20**	**0.10**	**0.10**	**0.10**
	[-.02; .12]	**[.03; .20]**	**[.07; .21]**	[-.16; .18]	**[-.39; -.03]**	**[.04; .36]**	**[.04; .16]**	**[.02; .17]**	**[.04; .17]**
IND	-0.01	-0.03	0.05	**-0.28**	**-0.21**	-0.11	0.09	0.09	**0.13**
	[-.10; .08]	[-.13; .08]	[-.04; .14]	**[-.45; -.12]**	**[-.39; -.02]**	[-.27; .04]	[-.04; .21]	[-.06; .25]	**[.01; .26]**
CON	0.04	0.02	0.08	**-0.32**	**-0.29**	-0.09	**0.13**	0.08	**0.15**
	[-.05; .13]	[-.08; .13]	[-.05; .10]	**[-.55; -.08]**	**[-.55; -.01]**	[-.31; .13]	**[.04; .22]**	[-.03; .19]	**[.06; .24]**
MIN	0.05	0.07	**0.08**	**0.10**	0.07	0.04	-0.04	-0.02	**-0.06**
	[-.01; .10]	[.00; .13]	**[.03; .13]**	**[.00; .21]**	[-0.06; .19]	[-.06; .14]	[-.10; .02]	[-.08; .05]	**[-.12; -.00]**
Between students (level two)									
AGE	-0.03	**0.09**	-0.04	0.02	0.08	-0.02	-0.06	0.03	-0.06
	[-.09; .04]	**[.02; .16]**	[-.12; .03]	[-.04; .14]	[-.01; .17]	[-.10; .07]	[-.12; .02]	[-.03; .10]	[-.13; .02]
GEN	0.05	**0.10**	0.04	0.10	0.08	0.03	**0.08**	**0.09**	0.04
	[-.03; .13]	**[.02; .19]**	[-.05; .12]	[-.02; .21]	[-.04; .19]	[-.09; .14]	**[.01; .16]**	**[.01; .17]**	[-.04; .13]

Note. TEX = teacher explanations; GWO = group work; DIS = whole-class discussions; IND = individual seatwork; CON = copying notes; MIN length of lesson expressed in minutes. AGE = age of student; GEN = gender of student. Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits: English: Sample size = 702; PPP = 0.314, *ƒ* (95% CI) = -48.60 - 72.93; Kinyarwanda: Sample size = 374; PPP = 0.307, *ƒ* (95% CI) = -44.30 - 78.87; mathematics: Sample size = 773; PPP = 0.157, *ƒ* (95% CI) = -32.35 - 96.33. Coefficients in bold are significantly different from zero.

To what extent are Rwandan students' lesson-specific experiences of expectancy, utility value, and cost influenced by classroom activities?

Table 44 on the previous page presents the standardised within-level effects of classroom activities on motivational states (see β_t, β_d, β_p, β_i, and β_n in figure 16). As table 44 highlights, the nature of motivational effects associated with classroom activities varied strongly across school subjects. We therefore felt incapable of accepting any of the hypotheses 2a through 2c for all subjects. Within subjects, it was slightly different. Those activities we associated with student-centred pedagogies (GWO = group work; DIS = whole-class discussions in table 44) exerted positive and significant effects especially in mathematics but also in English. In mathematics both group work and whole-class discussions positively predicted especially students' competence and cost (self-) evaluations but also utility perceptions. We therefore considered hypothesis 2a supported by the data only in the case of mathematics. Given that group work and whole-class discussions were the least common classroom activities within mathematics lessons (see table 41), maths teachers may be best advised to make more use of both during lessons.

Within English, we only observed whole-class discussions to exert positive effects on motivation. Group work did not predict motivation at all, something we also observed within Kinyarwanda. In the case of languages, our study results contradicted the motivational literature, which implied positive effects of group work on competence (Bandura, 2000) and cost (self-) evaluations (F. Kirschner et al., 2009). Our results might have also pointed to potential inability especially of Rwandan language teachers to ensure group work leads to quality discussions amongst students (Webb, 2009). Within languages, group work often implies tasks that do not require to find the right solution but instead to develop interpretations and argumentations based on a text. To ensure students achieve this goal, teachers must provide clear instructions to the class on how to successfully complete group tasks. This may be particularly true in Rwanda, where students are often just not used to these kinds of tasks (Honeyman, 2015b). Language teachers in particular may therefore benefit from training interventions to build up their facilitation skills vis-à-vis whole-class discussions (English) and group work (English and Kinyarwanda). Within mathematics, this problem may be less severe because here teachers tend to give mathematics problems and exercises to groups expecting students to work together to find the right answer, something more aligned with Rwandan students' expectations (Honeyman, 2015b).

The motivational effects of those activities we associated with teacher-dominant activities (IND = individual seatwork; CON= copying notes in table 44) varied across subjects as well. Within Kinyarwanda, they exerted negative effects especially on competence and utility (self-) perceptions while in mathematics they were largely associated with positive effects on competence and cost (self-) perceptions. Within English, we did not observe any significant effects of teacher-dominant activities at all. Hence, we considered hypothesis 2b supported only in the case of Kinyarwanda. The positive motivational effects of these activities within mathematics were unexpected but they may be explained through the general perceptions of the purpose of schooling in Rwanda. As Honeyman (2015b) argues, Rwandan students do not expect their teachers to help them develop creativity or problem-solving skills. Instead, they expect schools to predominantly prepare them for national examinations they have to pass every three years. The provision of notes and exercises students can review at home as preparations for exams and examinations may then be one way to deliver on this expectation. It may be for this reason that classroom activities, otherwise described as "autonomy supressing" (Assor et al., 2002), can indeed be associated with higher motivational states. Interestingly, all classroom activities had positive effects on students' mathematics-related cost perceptions. Whatever the type of activity, students seemed to find just studying in class helpful for copying with the challenges associated with mathematics.

The motivational effects of teacher explanations (TEX in table 44) varied across subjects, too. Whereas explanations hardly exhibited any effects in Kinyarwanda and mathematics, they positively influenced students' competence, utility, and cost (self-) perceptions in English. In Rwanda, especially basic-education students, who make up the majority of our sample, oftentimes exhibit just poor command over English despite it being the medium of instruction (Williams, 2019). Michener et al. (2017) found teacher explanations to predict second-language acquisition. Indeed, teacher explanations may be those moments during English lessons at Rwandan schools where teachers provide Kinyarwanda translations for new vocabulary, explain grammar or point out the meaning of texts studied. This in turn may help Rwandan students to develop more confidence about the language, to perceive it as less challenging, and more helpful for their own (school) life.

We rerun all three subject-specific models while controlling for students' lesson-specific novelty perceptions. Across subjects, novelty turned out to be a fairly consistent predictor of motivational states. Across English, Kinyarwanda, and mathematics, it positively predicted competence and utility (self-) perceptions and tended to negatively predict cost evaluations. Perceived novelty has been associated with situational interest, which in turn has been found to be positively related to competence self-perceptions (Linnenbrink-Garcia et al., 2013) and utility value (Hulleman, Godes, et al., 2010). That perceived novelty is negatively related to cost perceptions suggest that the introduction of a new topic may be associated with additional work amongst students. Also, controlling for novelty rendered almost all effects of teacher explanations insignificant. This is not surprising given that both perceived novelty and the time dedicated to teacher explanations is correlated. It suggests that the motivational effects of teacher explanations were somewhat mediated by the extent to which they are used to introduce new learning topics to students.

The effects of age and gender on lesson-specific motivational states

As highlighted by table 44, neither age (AGE) nor gender (GEN) was an important predictor of students' average motivational states that emerged at level 2. Only average utility perceptions in English were predicted by age and gender while average competence and utility states in mathematics were positively predicted by gender. We therefore concluded hypothesis 3 as refuted given our data. This is in contrast to Bulla et al. (2022b). Using school subject-specific (self-) evaluation measures, they found age and gender to predict motivational self-reports at least partly. It suggests that average motivational states and school subject-specific self-reports may potentially be distinct constructs (also see Bulla et al., 2022c). As for the additional control variables at level 1, we found that the length of lessons expressed in minutes hardly predicted motivational states of students. Day of the week only seemed to matter within mathematics. Mondays seemed to be associated with slightly less favourable motivational states. Finally, time of the day only seemed to be relevant within Kinyarwanda. Kinyarwanda lessons in the afternoon were associated with slightly higher motivational levels.

Conclusions

In 2016, Rwanda adopted the student-centred pedagogy for its classroom teaching. According to our observations, teacher-dominant activities in form of copying notes and individual seatwork were still the most common types of lesson activities in Rwandan classrooms. Nonetheless, student-centred activities were not that uncommon either with the vast majority of lessons observed having at least some time dedicated to student-centred tasks. One could expect that over time Rwandan classrooms become more student-centred. After all, educational reforms such as the introduction of student-centred pedagogies are resource heavy (Schweisfurth, 2011) and thus difficult to shoulder for low-income countries such as Rwanda. Over time, this situation may improve.

In terms of the motivational effects of classroom activities, we did not find consistent patterns across school subjects. We found student-centred activities to have positive effects mainly in mathematics while teacher-dominant activities exerted negative effects especially in Kinyarwanda. Within English, teacher explanations were the main predictors of students' motivational self-reports. These patterns suggest that Rwandan students may have different "motivational needs" across subjects. Thus, schools and teachers would be best advised to adopt school subject-specific lesson planning rather than imposing a general teaching model across all subjects. Future studies on the motivational effects of activities within Rwandan classrooms need to further explore this.

5.6. Manuscript five: Is being on-task predicted by motivational self-reports and classroom activities? Results from an observational study in Rwanda

Abstract

Student-centred classroom activities are considered to be conducive to student learning. Behavioural engagement, a product of motivation, is considered a prerequisite of learning. Within Africa, the effects of classroom activities and student's motivation on engagement have not yet been studied. Within the present study from Rwanda, we related 1620 lesson-specific motivational self-reports across 263 lessons in English, mathematics, and Kinyarwanda provided by 188 lower-secondary students (mean age = 15.1 standard error = 1.62; 56.4% female) with in-situ observations of teaching activities these students encountered and the subsequent engagement levels they exhibited during class. Three-level modelling revealed student-centred classroom activities did and motivational self-reports did not predict engagement. The latter may be due to the way the measures were operationalised.

Keywords: Student-centred pedagogy; behavioural engagement; student motivation; classroom observations; multilevel modelling; Rwanda

Introduction

Educational systems in many parts of Africa are characterized by poor learning outcomes amongst students (Sandefur, 2018). Many different reasons have been identified to explain poor learning outcomes in Africa and other parts of the so-called developing world. Within these debates, however, the role of behavioural engagement that students exhibit in class appears to be hardly reflected on. This is surprising because engagement is “the direct (and only) pathway” (Skinner & Pitzer, 2012, p. 23) to attain learning. As a contribution to address the learning crises in many parts of Africa and beyond, it is important to better understand the determinants of student’s behavioural engagement in class. Studies have already examined the connection between engagement as well as socio-economic status (e.g., Ansong et al., 2018), and teacher support (e.g., Ansong et al., 2017). Two other sets of correlates, however, appear to be missing from educational research on engagement within African contexts. First, Africa-based research on the nexus between engagement and motivation appears to be largely missing, which is surprising because engagement has been dubbed as “the outward manifestation of motivation” (Skinner et al., 2009, p. 224). In line with expectancy-value theory (e.g., Eccles & Wigfield, 2020), students’ competence, utility, and cost (self-) evaluations should thus predict the level of engagement students exhibit in class. A second set concerns contextual factors in form of classroom activities. Within the international-development discourse on education, classroom practices have often been discussed using the of student-centred versus teacher-centred pedagogies (e.g., Frost & Little, 2014; O’Sullivan, 2004; Sakata, 2022; van de Kuilen et al., 2019). The former is considered superior over the latter because they are assumed to be more effective in helping students to learn (Schweisfurth, 2013). Given that engagement is the only pathway to learning, any effects of classroom activities on learning outcomes must be mediated by engagement.

Within African contexts, however, the effects of student-centred activities within classrooms on engagement within classrooms have hardly been studied if at all (see, for example, Bremner et al., 2022). Within the present study, we employed an intensive-longitudinal design to investigate the relations of both motivational and contextual factors with engagement levels exhibited in class.

Behavioural engagement as key to learning

Student engagement is defined as a "student's active participation in academic and cocurricular or school-related activities, and commitment to educational goals and learning" (Christenson et al., 2012, p. 816). Engagement has been conceptualised as multidimensional construct that manifests itself at different levels. It manifests itself at different levels because student may engage at the school level, the classroom level or even the task level (Skinner & Pitzer, 2012). Engagement at the classroom level refers to the extent to which students are focussed on lesson activities. Especially at the level where students encounter concrete academic tasks and challenges, engagement has been conceptualised multidimensionally in the sense that it combines a behavioural, cognitive, and emotional dimension (Fredricks et al., 2004). Compared to cognitive and emotional engagement, behavioural engagement appears to be a stronger predictor of academic achievement (Lei et al., 2018). Behavioural engagement may also constitute a necessary requirement for students to get involved in academic activity cognitively and emotionally. For example, only if I allow myself to work on a task, I then can start experiencing joy about it. At the task level, behavioural engagement constitutes a process of states. States refer to the behaviours, thoughts, and feelings that are exhibited at a particular time (Baumert et al., 2017). As a state, behavioural engagement has been conceptualised to fluctuate between being off and on task. Being behaviourally disengaged in form of being off-task implies “any behaviour that does not involve the learning task or material, or where learning from the material is not the primary goal” (Rodrigo et al., 2013, p. 2). By contrast, being behaviourally engaged in form of being on-task refers to a “state in which attention is directed more or less exclusively to one target or task” (Ruff & Rothbart, 2001, p. 110). Hence the state of heightened attention which is implied by being behaviourally on task may explain why behavioural engagement has been found to be a strong predictor of learning (Connor et al., 2009; Fisher et al., 2014; Gobel, 2008; Godwin et al., 2021; Guillaume & Khachikian, 2011; Kupiainen et al., 2014; Louw et al., 2008; Moffett & Morrison, 2020). Based on theoretical and empirical work, Hospel et al. (2016) further differentiated the behavioural states associated with engagement. They sub-divided behavioural disengagement into being passively (“withdrawal”) and actively off-task (“disruptive behaviours”). Similarly, they sub-divided behavioural engagement into being passively (“following instructions”) and verbally on-task (“participation”). These different categorizations of engagement are found within current observation protocols on measuring behavioural engagement (Ocumpaugh et al., 2015; Shapiro, 2011). These different categories are not stable person characteristics. Engagement states are rather short-lived phenomena since they are the product of a person’s stable characteristics (i.e., traits) and those contextual factors that make up a specific situation (Geiser et al., 2013). Applied to the case of behavioural engagement, this implies that on-task states may vary between (e.g., Vasalampi et al., 2020) and even within lessons (e.g., Salmela-Aro et al., 2021). Studies on the extent to which students exhibit these different behavioural states within African school contexts have not been reported yet. As discussed next, classroom activities may be an important determinant of students’ level of task engagement.

Student-centred classroom activities as contextual predictors of engagement

Classroom activities constitute those "opportunities to learn" (Jansen, Scherer, et al., 2015, p. 15) students engage with during a lesson. Within the policy discourse in Africa, classroom activities are often discussed using the lens of student-centred versus teacher-centred practices (e.g., Frost & Little, 2014; O'Sullivan, 2004; Sakata, 2022; van de Kuilen et al., 2019). Student-centred approaches are centred around the idea of letting students steer their own learning, which allows them to develop and construct knowledge. The role of teachers is mainly facilitatory in nature in the sense that they help to keep learning processes in line with students' prior knowledge and current needs (Bremner, 2021). One of the key principles to ensure student centredness of classroom activities has called "interaction" (Bremner, 2021, p. 166), "dialogue" (Schweisfurth, 2013, p. 146) or "active teaching" (Ngware et al., 2012). It refers to the creation of opportunities that allow students to verbally interact not only with the teacher but also with each other (Bremner, 2021). Verbal interactions within classrooms may appear especially in form of whole-class discussions and group work. The former can be loosely defined as "all verbal exchanges where one individual addresses another individual or individuals and at least one addressed individual replies" (Howe & Abedin, 2013, p. 325). Small-group work implies students being divided into groups where they work together to complete a given task through verbal interactions (e.g., Murphy et al., 2009). Using observations to gauge on- and off-task states, Nguyen et al. (2016) have found both student-student as well as student-teacher interactions to predict behavioural engagement.

Within teacher-centred activities, it is predominantly the teacher who controls the pace and content of the learning process. The most eminent example in this respect may be whole-class explanations provided by the teacher on concepts, ideas, and procedures that students need to internalize in order to meet learning goals. Unlike classroom discussions or group work, guided instructions constitute “one-way teacher-to-learner transmission” (Schweisfurth, 2013, p. 147). According to the student-centred lens, this would curtail students’ “autonomy” (Bremner, 2021, p. 172) because it prevents them discovering essential information for themselves. According to the direct-instruction paradigm, however, providing clear and guiding explanations is more effective especially when dealing with young learners and novel information (P. A. Kirschner et al., 2006). Olivier et al. (2020) used self-reports to gauge engagement and found it to be correlated with measures that are loosely comparable to student-centred practices (“autonomy support”) as well as explanations (“structure”) given by teachers (also see Skinner & Belmont, 1993).Within the international-development discourse, however, the emphasis on student-centred teaching might have been especially a response to the traditional way of teaching that continue to be common within low- and middle-income countries (Nag, 2014; Schweisfurth, 2013) and that are associated especially with surface (rote) learning (Nag, 2014). Within classrooms, this often manifests itself in form of students copying large amounts of notes from the blackboard and then to practice what they copied through a series of individual seatwork exercises (Frost & Little, 2014; Honeyman, 2015b; Sakata, 2022). They constitute “autonomy-suppressing” (Assor et al., 2002) types of classroom activities because they are administered to all students in class irrespective their individual needs, requirements or interest. One should therefore expect that the extensive use of note copying and individual seatwork in class may increasingly lead to forms of off-task behaviour amongst students.

Competence, utility, and cost (self-) evaluations as motivational predictors of engagement

Classroom activities do not directly shape students' engagement levels. Students' subjective perceptions of classroom activities mediate their effects on engagement (e.g., Olivier et al., 2020). The formation of these subjective perceptions is discussed within the context of achievement motivation, which refers to those psychological (proximal) processes that directly shape student actions exhibited within academic-performance contexts (Wigfield et al., 2006, 2015). According to the expectancy-value theory (e.g., Eccles & Wigfield, 2020), task behaviour is shaped by one's expectations to successfully complete the task (success expectancies), the value one places on that task (subjective task value), and by how taxing the completion of that task is perceived to be (cost evaluations). Conceptually, success expectancies are assumed to be shaped by self-evaluations of one's academic performance in the past (i.e., "self-concept of ability"; Eccles et al., 1983; Wigfield & Eccles, 2000) but empirically both belief types often turned out to be inseparable factors (e.g., Eccles, Wigfield, et al., 1993). Within the present study, we thus use them interchangeably (e.g., Wigfield & Gladstone, 2019) and refer to them as competence self-evaluations. Subjective task value is conceptualised multidimensionally (Eccles & Wigfield, 2020). One of its dimensions is utility value, which describes how useful a learning task is perceived to be for one's life outside school (Eccles & Wigfield, 2020). Compared to other value dimensions such as intrinsic or attainment value, utility value has been found to be malleable through interventions (e.g., Harackiewicz et al., 2016; e.g., Hulleman, Godes, et al., 2010). This property makes utility value of educational interest and is thus the value aspect we focused on within the present study. Task-specific cost perceptions are multidimensional, too. One of its dimensions concerns perceived effort costs. It refers to the amount of work and attention the completion of a task requires (Flake et al., 2015). In line with predictions, studies found that competence (e.g., Dietrich et al., 2017; Olivier et al., 2020; Salmela-Aro et al., 2021), and value perceptions (e.g., Dietrich et al., 2017; Olivier et al., 2020; Salmela-Aro et al., 2021) positively predict engagement whereas unfavourable cost evaluations (e.g., Dietrich et al., 2017) is a negative predictor of engagement. Respective relations appear to not have been studied within the African context. To address this research gap, we carried out a study in Rwanda, East Africa.

The Rwandan School Context

Rwanda, East Africa, is a low-income country with Kinyarwanda as its national language (NISR, 2016). As part of 12 years of free basic education (12YBE), students first attend six years of primary schooling before going for three years of lower (S1 – S3) and then upper secondary education (S4 – S6). Since 2016, Rwanda's Ministry of Education has also been promoting student-centred approaches (van de Kuilen et al., 2020), which include "participatory and interactive methods that engage young people in the learning process" (MINEDUC, 2015, p. 63). This should have opened up space within classrooms for students to be engaged not only passively (e.g., copying notes) but also verbally (e.g., participation classroom discussions). Two factors might have hampered the general practice of student-centred pedagogies across Rwandan classrooms. First, culturally Rwanda is organized hierarchically (Honeyman, 2015b; E. King, 2014) something observed within the education domain as well. The teacher has traditionally been considered "the central arbiter of knowledge" (p. 161) and therefore classroom learning was mainly organized around "teachers teach; students listen" (Honeyman, 2015b, p. 136), something not uncommon for many African contexts (e.g., B. Fuller & Snyder, 1991; Stambach, 1994). These culturally ingrained ways of teaching and learning are difficult to abandon and thus any top-down adoption of new pedagogies may take time until they unfold any effects in classrooms. Second, across secondary schools students often exhibit only poor language skills of English (Honeyman, 2015b; Williams, 2019). It might have been one of the ramifications of the shift of the medium of instruction from French to English in 2009 that was executed with just a short preparatory phase for all teachers (Abbott et al., 2015; Honeyman, 2015a). As a result, secondary students often significantly lag behind in terms of those language standards in English they should have attained given their class levels (Burdett & James, 2018; Milligan et al., 2016). Without some level proficiency in English, it is hard to engage verbally in class. As a result, at least until some years ago, students might have focussed especially on copying notes (Honeyman, 2015b) or working their exercises silently (Milligan et al., 2016). More recent evidence suggests that things might have changed. Based on an observational study using a sample 12 lessons, van de Kuilen et al. (2020) found student-centred activities to be incorporated into the teaching. They observed instances of "whole class questioning during which the teacher, with the help of hints and clues, enabled learners to discover new knowledge themselves" (p.11). Hence, additional research is warranted to explore to what extent Rwandan students engage in class not only passively but also verbally. To our knowledge, no study has yet investigated the relations between classroom-based engagement and motivation amongst Rwandan students.

The present study

Within the present study, we focussed on classroom-based behavioural engagement as exhibited by Rwandan secondary students within the subjects of English, Kinyarwanda, and mathematics. In particular, we were interested in its scope as well as how it relates to classroom activities as well as students' experiences of competence, utility, and cost. Unlike motivation, engagement is only partly a latent construct because it describes "the *observable* qualities of students' actions with academic tasks" (Skinner & Pitzer, 2012, p. 24; emphasis added). This in turn makes it possible to gauge students' level of behavioural engagement through observing their behaviours towards this task. In this respect, a student's gaze and head orientation have been found to be strong predictors of attention (e.g., Stiefelhagen, 2002). Instead of employing expensive hardware within Rwandan classrooms, we gauged states of behavioural engagement through systematic observations (Salvia et al., 2007; Sanders et al., 2010; Wragg, 2012) carried out by external raters located in class (Godwin et al., 2016, for studies wth a similar observation design see 2013). This approach is effective especially when the focus is on frequency rather than the quality of "specific behaviours" (Sanders et al., 2010, p. 425). The same also applies to classroom activities. In particular the type of activity can be easily determined through external observations. Being latent variables, motivational states are generally measured through self-reports (Appleton et al., 2006). Signalling devises would allow collecting motivational self-reports around the time when engagement levels are gauged. Within low-resource contexts such as public schooling in Rwanda, however, they might have attracted too much of curiosity, when operated in classes, and potentially jealousy, when handed out only to some students. It was therefore deemed not an option to collect motivational self-reports through signalling devises. Instead, they were collected upon completion of each lesson were carried out observations (i.e., event-contigent design; e.g., Bolger & Laurenceau, 2013). This design was not in line with the conceptualisation of motivation being an "antecedent" of engagement, but we considered it the only viable option. The study was guided by the following research questions and hypotheses.

1) *To what extent is Rwandan students' behavioural engagement subject to intrapersonal variability both within and between lessons?* In line with the literature, we expected

behavioural engagement to exhibit substantial variability not only between students but also between lessons as well as within lessons. Moreover, we expected the biggest share of variability to originate at the within level (hypothesis 1).

2) *To what extent do classroom activities predict Rwandan students' behavioural engagement in class?* Being a "situational" variable, we expected behavioural engagement exhibited during lessons to be influenced by contextual factors. Within lessons, we considered classroom activities to be major sources of contextual influences on engagement. In line with the literature, we expected those activities associated with student-centred practices to exert positive influences on engagement (hypothesis 2a) while those activities we linked with teacher-dominant practices were expected to exert negative effects (hypothesis 2b). Teacher explanations were also assumed to be positively related to engagement levels (hypothesis 2c).
3) *To what extent is Rwandan students' behavioural engagement in class related to their lesson-specific motivational self-reports?* Being the "outward manifestation of motivation", we assumed engagement to be positively predicted by students' competence, utility, and cost self-reports. We expected competence but also cost (self-) evaluations to be stronger predictors of engagement than utility evaluations (hypothesis 3).

Method

Procedures

Data reported here was collected as part of a research project on achievement motivation within Rwanda (e.g., Bulla, Dominik et al., 2022; Bulla et al., 2022b, 2022a). Classroom observations and motivational self-reports were collected repeatedly within the subjects of English, Kinyarwanda, and mathematics for 13 consecutive days at the beginning of the 2020 Rwanda school year. Lower-secondary students (secondary 1 through secondary 3) were enrolled at three public schools outside Kigali (see table 45), two basic-education (A2 and B1) and one boarding school (B2). In Rwanda, boarding schools were traditionally the providers of secondary education whereas basic-education schools were introduced to absorb the increasing number of student enrolments post introduction of 12YBE (Williams, 2019). For the present analysis, we used observational and self-report data collected across 263 (i.e., 93 English, 46 Kinyarwanda, and 124 mathematics) lessons (see table 46 on the page). We had removed those lessons were students only copied notes (12 lessons) or were sitting exams (7 lessons). We also removed those cases where we only had less than three engagement ratings per lesson. Within each school subject-specific dataset we also excluded those cases for whom we recorded data on not more than two different lessons. As table 46 highlights, we did not collect any data on Kinyarwanda at the boarding school B2. The table also shows that Kinyarwanda lessons appeared to be slightly shorter than English and mathematics lessons expressed in minutes.[29] This also means we collected fewer observations within Kinyarwanda lessons.

Table 45: School sample characteristics (manuscript five)

Characteristics	A2	B1	B2
Boarding school?	No	No	Yes
N (% female)	39 (66.67)	93 (61.29)	56 (41.07)
Average class size (se)	60.6 (9.11)	48.55 (7.02)	43.2 (8.55)
Age (se)	15.31 (1.59)	15.51 (1.52)	14.17 (1.48)

Note. ‘N’ refers to sample size. ‘se’ refers to standard errors.

[29] In Rwanda, a lesson period is supposed to take 40 minutes. In case of double or triple lessons, however, the length of a lesson can be longer than 40 minutes. Also, teachers did not always start or ended lessons on time, which added additional variability to average lesson lengths.

Across all subjects, we collected 16,986 (i.e., 6,318 in English, 1,840 in Kinyarwanda, and 8,828 in mathematics) engagement ratings and the same number of classroom-activity ratings. Across the 263 lessons, 1620 (i.e., 596 in English, 245 in Kinyarwanda, and 779 in mathematics) lesson-specific self-reports were collected. All self-reports were collected in the last minutes of a lesson once the teacher had wrapped up the teaching. The self-report questionnaires, in Kinyarwanda only, always elicited personal information (i.e., name, age, gender, class), which was then used to track student responses across lessons. Data sets were imbalanced not only between subjects but also between students within each subject. It was 4.25 ($standard\ error = 1.20$) in English, 3.60 ($standard\ error = 0.67$) in Kinyarwanda, and 4.81 ($SD = 1.52$) in mathematics. All questionnaires were entered independently by two separate data-entry clerks. Mismatches during data entry between both clerks were resolved by reverting back to the hard copies of the questionnaires. For students to partake in the study, they were asked to sign a child assent form. Head-teacher consent was sought as they acted *in loco parentis* (British Psychological Society, 1992). As a token of appreciation, all students that submitted consent forms participated in a tombola carried out after field work. Per class, three students won a pocket calculator worth around 75p. The study has ethics approval from University of Oxford (ED-CIA-19-084) and University of Rwanda (266/CMHS IRB/2018).

Table 46: Distribution of lessons attended during data collection (manuscript five)

School	English	Kinyarwanda	Mathematics	Total/ average
A2	21	12	42	75
B1	49	34	52	135
B2	23	0	30	53
Total	93	46	124	263
Minutes (se)	52.3 (19.5)	39.9 (19.8)	55.9 (24.0)	51.8 (22.5)
Ratings (se)	67.9 (36.3)	40.0 (20.6)	71.2 (35.7)	64.6 (35.5)

Note. ‘Minutes’ refers to the length of lessons expressed in minutes. ‘se’ refers to standard errors.

Participants

Across the three schools, we included 22 lower-secondary classes into the study. Two classes were dropped. With student numbers of almost 100 students each they were just too big to carry out classroom observations effectively. Student observations did not target all students in class. Instead, for each class we randomly selected 10 students from class register. Not all students selected were found in call all the time. Given the restrictions of at three six self-reports, we ended up using data from 188 students (140 in English; 68 for Kinyarwanda; 162 in mathematics). Only 30 provided at least three self-reports within each of the three subjects. The full sample was distributed across 22 classrooms across the three schools. The average number of study participants per class was 8.55 ($SD = 1.43$) and ranged from 5 to 10 students. Table 45 provides additional information on sample characteristics. Students from the boarding school B2 were less likely to be female as well as younger compared to basic-education students.

Table 47: The lesson-specific motivational scales (manuscript five)

	Item
Competence scale	
LC1	Do you think you understood today's activities of [SUBJECT]?
LC2	Do you think you were capable to do today's activities of [SUBJECT]?
LC3	Do you think you understood today's activities of [SUBJECT]?
Daily-utility scale	
DU1	Do you think you will use what you learned in today's activities in [subject] in your everyday life?
DU2	Do you think today's activities of [SUBJECT] will help you in your daily life?
DU3	Do you think today's activities of [subject] will be useful for your everyday life?
Cost scale	
LE1	Were today's activities of [SUBJECT] hard?
LE2	Were today's activities of [SUBJECT] tiring?
LE3	Were today's activities of [SUBJECT] confusing?

Measures

Lesson-specific self-reports

Instead of translating lesson-specific motivational scales into Kinyarwanda, we developed them for the Rwandan school context. The scale-development process largely followed the one for school subject-specific motivational beliefs outlined in Bulla et al. (2022b). These scales were found to exhibit predictive validity vis-à-vis school grades (Bulla et al., 2022c) as well as classroom activities aggregated at the lesson-specific level (Bulla et al., 2022a). Lesson-specific competence, daily utility, and cost perceptions were each measured through three items (see table 47). Students responded to each item using a three-options response format consisting of 1) yes, very much; 2) A bit; or 3) no, not at all. All items were coded in such a way that higher values represent more favourable perceptions. Table 48 presents average scores for each scale, subject, and scale.

Table 48: Average score per scale (manuscript five)

Scale	School			
	A2	B1	B2	All
English				
Comp	2.39 (0.5)	2.58 (0.46)	2.69 (0.43)	2.58 (0.47)
Daily	4.42 (0.89)	4.02 (1.06)	3.67 (0.96)	3.98 (1.03)
Costs	2.24 (0.55)	2.37 (0.54)	2.44 (0.55)	2.37 (0.55)
Kinyarwanda				
Comp	2.4 (0.49)	2.78 (0.39)	NA	2.71 (0.44)
Daily	3.79 (0.92)	4.38 (0.89)	NA	4.26 (0.93)
Costs	2.26 (0.54)	2.51 (0.52)	NA	2.46 (0.53)
Mathematics				
Comp	2.4 (0.52)	2.46 (0.53)	2.53 (0.51)	2.47 (0.52)
Daily	3.76 (0.97)	4.06 (1.07)	3.72 (1.01)	3.89 (1.04)
Costs	2.23 (0.51)	2.21 (0.58)	2.23 (0.53)	2.22 (0.55)

Note. ‘Basic’ = basic-education school; ‘Boarding’ = boarding school. Numbers in brackets refers to standard errors associated with the scale scores.

Compared to basic-education students, boarding-school students appear to have slightly higher average scores in case of English competence and slightly lower scores in case of English utility perceptions. For each of the three school subjects, table 49 presents within-level polychoric correlations for lesson-specific scales and single-level polychoric correlations for domain-specific motivational scales as measures of instrument reliabilities. Correlations were substantial at the within-person level but particularly high, and partly perfect, at the between-person level. Together, we took this as evidence of acceptable scale reliabilities.

Table 49: Average within-scale polychoric correlations (manuscript five)

Scale	English	Kinyarwanda	Mathematics	Average
Average within-level polychoric correlations (lesson-specific self-reports)				
Competence	0.49	0.56	0.65	0.57
Daily utility	0.49	0.71	0.63	0.61
Cost	0.55	0.41	0.5	0.49
Average	0.51	0.56	0.59	
Average between-level polychoric correlations (person-specific self-reports)				
Competence	0.98	0.92	0.95	0.95
Daily utility	0.96	0.92	0.97	0.95
Cost	1	1	0.97	0.99
Average	0.98	0.95	0.96	

Note. Polychoric correlations were estimated using a multilevel framework. Level one represented the lesson-specific within-person level and was nested within level two which constituted the between-person level. The correlational analyses concerning the lesson-specific measures were performed using the ‘TYPE = TWOLEVEL BASIC’ command.

Classroom observations

Student engagement levels and classroom activities were rated using the coding schemes in table 50 on page 308. A lesson was observed by one rater. The coding schemes were derived from the literature and field tested prior to the study. Behavioural engagement was operationalised through being ‘on task’ as well as being ‘verbally on task’. Behavioural disengagement was operationalised through being ‘off task’ and being ‘disruptively off task’

(Ocumpaugh et al., 2015; Shapiro, 2004, 2011). On- or off-task was determined especially through the student's gaze and head orientation (Godwin et al., 2016). During observations, raters determined solely the nature of a student's behavioural state but not its duration, intensity or quality (Alberto & Troutman, 2013). Table 51 presents the frequencies of the different behavioural states observed. Given the low frequencies of verbal engagement and disruptive disengagement, we collapsed behavioural engagement into a binary variable consisting of an off-task and on-task outcome. To code classroom activities, we adjusted the World Bank's (2015) revised version of Stallings' (e.g., 1977) Classroom Snapshot Observation System. The coding scheme consisted of 11 categories grouped into instructional, managerial, and miscellaneous activities. Within the analysis, we only focused on five activity types since the other six were hardly observed in class. Jointly, those five categories accounted for 90% (standard error = .30) of classroom activities observed. Observations to determine engagement and classroom activities followed a round-robin design (Ocumpaugh et al., 2015) organized around a 30-second long three-staged observation cycle. During the five seconds of the first stage, the rater observed and rated current lesson activities using the classroom-activity coding frame (see table 50). During the five seconds of the second stage, the rater then observed and rated the engagement level of a given focal student using the behavioural coding frame in table 50 while taking into account what behaviours can be expected given prevailing classroom activities (Godwin et al., 2016). Again, to avoid reactivity effects raters avoided direct eye contact and instead relied on side glances. During both coding stages, the classroom activity and student behaviour were chosen that accounted for the biggest proportion of the interval (i.e., predominant-activity sampling; Hutt & Hutt, 1970, p. 69). A multi-timer app guided raters through cycles. To prevent any order effects, the sequence in which to rate the 10 target students was always randomly determined ahead of the lesson. Ahead of lesson start, raters identified their location in class by reading out their names as well as the names of 10 additional, randomly selected students. The sole purpose of the latter group was to divert attention from the fact that observations were always targeting the same 10 students per class. Target students not found in class were marked as being absent. All raters completed a 15-day long training and were subjected to regular reliability checks. During data collection, the first author performed regular spot-checks of all raters, which implied parallel but independent classroom observations next to the rater concerned. This allowed determining average interrater reliability using Cohen's Kappa (1960). The average rates of agreements were substantial in case of both student engagement (.72 [.68; .76]) and classroom activities (0.80 [.77; .82]; Landis & Koch, 1977).

Statistical Analysis

The analytical strategy

Given the hierarchical nature of engagement with classroom observations (level 1) nested within lessons (level 2) and students (level 3), we specified multilevel models. To address the different research questions, we estimated a series of different models that all took into account the binary coding of engagement (e.g., Nguyen et al., 2016). To examine research question 1, we estimated a three-level unconditional mean model (e.g., Heck & Thomas, 2020) of engagement to calculate its intraclass correlations (ICCs). Within multilevel models of nominal data, there is no within-level variance. Conventional ICCs thus cannot be estimated. In modified formulas, the within-level variance is replaced by the variance of the standard logistic distribution ($\pi^2/3$), which in turn allows estimating for level-2 and level-3 ICCs for binary outcomes within 3-level models as follows (Sommet & Morselli, 2017): ($ICC_{level\ 2} = \frac{\sigma^2_{level\ 2}}{\sigma^2_{level\ 3} + \sigma^2_{level\ 2} + \pi^2/3}$; $ICC_{level\ 3} = \frac{\sigma^2_{level\ 3}}{\sigma^2_{level\ 3} + \sigma^2_{level\ 2} + \pi^2/3}$).

Knowing both estimates allows inferring the intrapersonal variability within lessons. We also estimated ICCs for the latent motivational variables but in these cases the unconditional model were two-level ones with lessons nested within students. To answer the second research question, we extended the initial unconditional mean model into a multilevel logistic model (e.g., Sommet & Morselli, 2017) by adding dummy variables representing the five different classroom activities at level 1 (i.e., conditional mean model; CMM) . Since the dummies only accounted for around 90% of classroom activities, we did not exclude any of them as reference variable. Being binary variables, we did not mean centre any of them. To address the final research question, we specified a multilevel structural equation model (ML SEM) that extended the previous conditional mean model by adding the motivational variables at level 2 (i.e., with a random-intercept parameterization; Heck & Thomas, 2020). Given that motivational self-reports were lesson-specific, we specified the model so that self-report items loaded at second and third level but not the within-level. In line with the assumption of cross-level measurement invariance (Stapleton et al., 2016), factor actor loadings were also constrained to be equal between levels. For both level 2 and level 3, the first factor loading was fixed to one. Both in the conditional mean as well as the structural equation model, we included a discrete-time variable as level-1, dummies representing day of the week as well as time of the day as level-2 and age, gender, school dummies as level3 covariates. The discrete-time variable represented the round at which an engagement state and the corresponding classroom activity was measured. Figure 17 on the next page presents a schematized overview of the multilevel structural equation model estimated to address the second research question.

Table 50: The classroom-activity and engagement coding scheme (manuscript five)

Activity	Definition
	Student-specific activity
Verbally On-Task	The student is not only paying attention to the classroom activity by having his/ her body and eyes oriented towards a relevant learning object but also is making (or wanting to make) verbal contributions to a lesson in form of answers to questions, clarifying questions, suggestions, etc.
On-Task	The student is paying attention to the classroom activity by having one's body oriented and one's eyes directed towards a relevant learning object (e.g., teaching talking, student talking, or blackboard with notes). A student may be active in that (s)he is coping notes or is working on a task. However, the student is not verbally active in that (s)he is asking any questions or is wanting to answer a question.
Off-Task	The student is not paying attention and thus his/ her eyes are directed towards anything but the learning source. His/ her body may also be oriented away from the classroom activity. However, even though the student is not paying attention, (s)he is not disturbing the lesson either.
Off-Task Disruptive	The student is disrupting the lesson through chatting, laughing, making noise, and/ or any unnecessary movements. In addition, His/ her eyes and posture is may also be oriented away from the classroom activity.
	Classroom activities
Teacher explanations	The teacher is explaining or demonstrating to the class tasks, homework, or any other academic content with the intention to make students understand it.
Group work	All students in the class are divided into groups where they are expected to work on a group assignment. In most cases, students rearrange their seats in accordance with how they have been divided into groups. During group work, the teacher generally goes from group to group to provide support and guidance.
Discussions	A student is verbally engaging with the teacher or other students to ask questions, to answer questions, or to discuss academic content while the rest of the class is listening.
Copying notes	Students are copying notes that the teacher is writing on the blackboard. At times, the teacher may read out passages written on the board mainly to help students copy notes properly.
Individual work	All students in class are working on a given task by themselves. In some cases, a student is at the blackboard to work the task there, potentially for demonstration purposes.

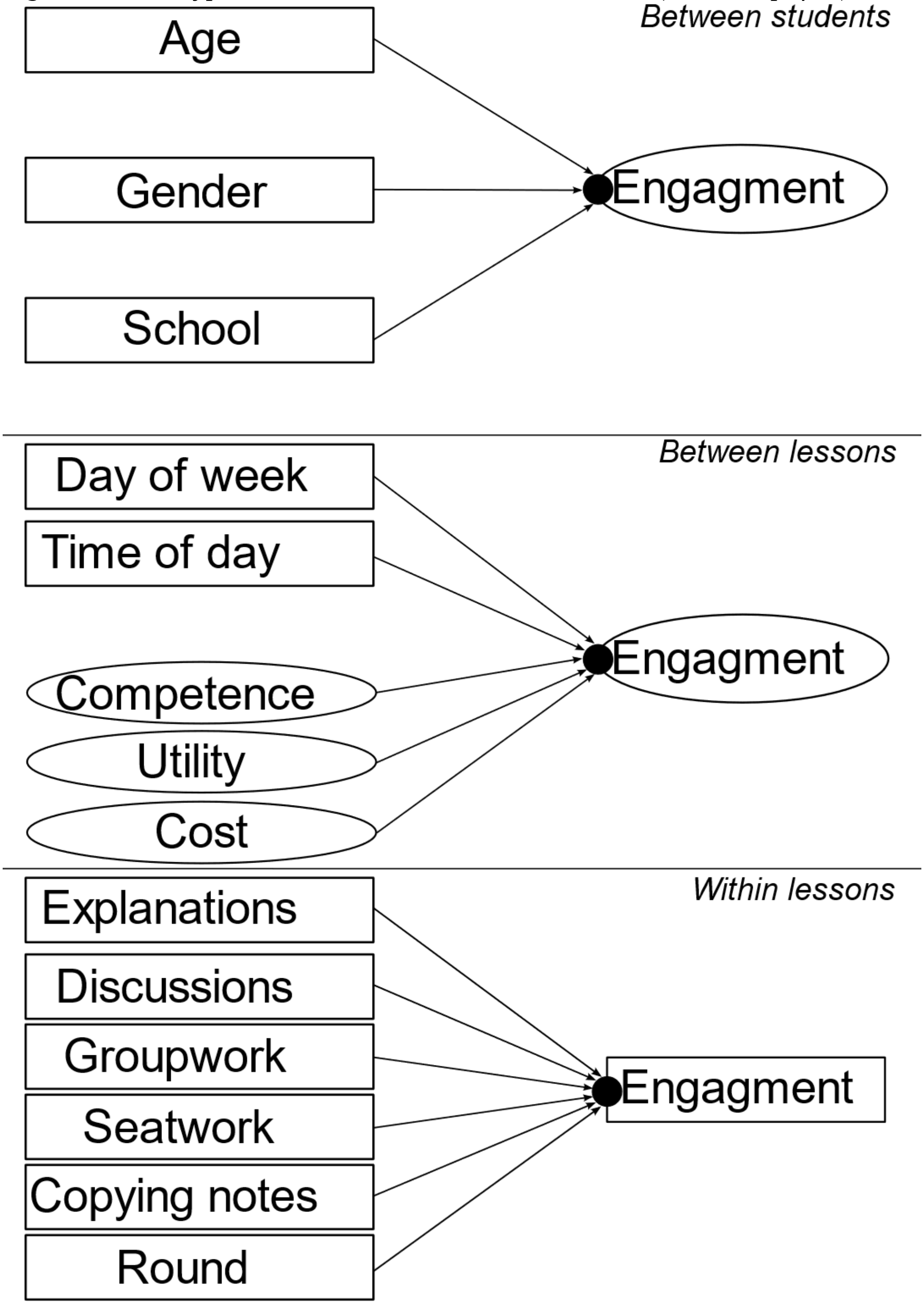
Note. The coding scheme on student engagement is adapted from Ocumpaugh et al. (2015) and Shapiro (2011). The coding scheme on classroom activities is adapted from World Bank (2015).

We run all models using the three-level Bayesian estimation framework in Mplus version 8.7 (L. K. Muthén & Muthén, 2017). The posterior distribution was obtained using the iterative Markov chain Monte Carlo (MCMC) algorithm (van de Schoot et al., 2014) using two chains and two processors used for parallel computing. As for the potential scale reduction (PSR) factor, we used the default option of .05. We discarded the first 1000 iterations as 'burn-in'. To reduce the effects of parameter autocorrelation between subsequent chain iterations, we only used every 30th iteration (thinning) during the computation of the final posterior distribution. We rerun every model with the number of iterations twice the number of iterations required during the first run. During all model estimations, we used Mplus' diffuse priors (for their hyperparameters see L. K. Muthén & Muthén, 2017). As part of the convergence diagnostics, we inspected the trace plots. We checked whether or not all chains converged to the same target distribution and whether or not all iterations were based on stable chains. Model fit itself was assessed through the posterior predictive p-value (PPP), which are computed using the likelihood-ratio chi-square statistic (Gelman, 2014). With one exception (i.e., the English-specific ML SEM), model fits generally turned out to be acceptable (PPP-values between .07 and .578, see table 53). We still considered all models as informative because in all cases, including the English-specific ML SEM, the credibility intervals contained the zero value.

Results and discussion

Table 51 presents the scope of engagement observed within lessons. As indicated earlier, verbal on-task behaviour was hardly observed in class not even during group work, let alone whole-class discussions. Given class sizes, it was no surprise not to observe target students being verbally engaged during classroom discussions. As table 45 highlights, on average there were up to 60 students in class on average. Given an average length of just 50 minutes per lesson (table 45), this would imply that a student would not even have one minute to make any verbal contribution. Based on a binary operationalization, which adds up both on-task categories as well as the two off-task categories, the average engagement level was 73% (SD = 44) and appeared to be equally distributed across schools and subjects. The only outlier might have been Kinyarwanda at basic-education B1 with a participation rate of 78% (SD = 34). As table 53 shows, by and large the round of observation when engagement was rated hardly seem to be an influence. Only within English it appeared that the longer into a lesson engagement was rated the higher the rating was.

Figure 17: The hypothesized three-level structural model (manuscript five)



To what extent is Rwandan students' behavioural engagement subject to intrapersonal variability both within and between lessons (hypothesis 1)?

For each of the three subjects, table 52 on page 303 presents level-2 and level-3 variances for the binary variable engagement and the per centages of the variances attributable to each of the two levels. To estimate ICCs, we used the variance of the standard logistic distribution as a proxy of the within-lesson variance. This parameterization resulted in rather low joint level-2 and level 3 ICCs ranging just from .068 (English) to .096 (Kinyarwanda). Within conventional settings, an ICC below .05 is used to justify a single-level parameterization of variables (B. O. Muthén & A., 1995). These low ICCs suggested that the bulk of the variance of engagement was accounted for by the within-lesson level ranging from .904 (Kinyarwanda) to .932% (English). We saw this to be in line with hypothesis 1. However, we did not expect the within-level to be so dominating, which we did not consider to be in line with the first hypothesis. It appeared that engagement states observed within class hardly depended on lesson or individual observed. This is contrast to number of studies who reported variances involving engagement that were substantially driven by between-student differences (A. J. Martin et al., 2015; Salmela-Aro et al., 2021; Vasalampi et al., 2020). For example, Martin et al. (2015) found around 65% and 56% of engagement and disengagement to be attributable to the between-student level. However, even though these studies employed intensive-longitudinal research designs to estimate ICCs, they also relied on self-reports to gauge engagement. These different methodological approaches might have partly explained the stark differences between our patterns and those reported in the literature.

Table 51: The extent of engagement and classroom activities observed (manuscript five)

School	Behavioural engagement				Classroom activities				
	Verbally on-task	Passively on-task	Passively off-task	Disruptively off-task	Teacher explanations	Classroom discussions	Group work	Individual Seatwork	Note copying
					English				
A2	0.01 (0.07)	0.72 (0.45)	0.24 (0.43)	0.03 (0.18)	0.19 (0.40)	0.14 (0.34)	0.10 (0.30)	0.18 (0.39)	0.24 (0.42)
B1	0.00 (0.06)	0.76 (0.43)	0.21 (0.41)	0.02 (0.14)	0.51 (0.50)	0.16 (0.36)	0.00 (0.03)	0.16 (0.36)	0.10 (0.30)
B2	0.01 (0.08)	0.70 (0.46)	0.26 (0.44)	0.04 (0.19)	0.20 (0.40)	0.19 (0.39)	0.05 (0.22)	0.30 (0.46)	0.17 (0.37)
					Kinyarwanda				
A2	0.00 (0.07)	0.71 (0.45)	0.25 (0.44)	0.03 (0.17)	0.35 (0.48)	0.18 (0.38)	0.08 (0.28)	0.12 (0.32)	0.17 (0.38)
B1	0.00 (0.00)	0.87 (0.34)	0.11 (0.32)	0.02 (0.14)	0.24 (0.43)	0.15 (0.36)	0.04 (0.20)	0.32 (0.47)	0.15 (0.36)
B2	NA	NA	NA	NA	NA	NA	NA	NA	NA
					Mathematics				
A2	0.00 (0.06)	0.75 (0.44)	0.22 (0.41)	0.03 (0.18)	0.29 (0.45)	0.11 (0.31)	0.10 (0.30)	0.26 (0.44)	0.12 (0.33)
B1	0.00 (0.03)	0.70 (0.46)	0.26 (0.44)	0.04 (0.19)	0.24 (0.43)	0.13 (0.34)	0.03 (0.16)	0.40 (0.49)	0.13 (0.33)
B2	0.01 (0.10)	0.73 (0.44)	0.23 (0.42)	0.03 (0.16)	0.25 (0.43)	0.09 (0.28)	0.12 (0.32)	0.40 (0.49)	0.10 (0.30)
					All subjects				
A2	0.00 (0.07)	0.73 (0.44)	0.23 (0.42)	0.03 (0.18)	0.27 (0.44)	0.13 (0.34)	0.10 (0.30)	0.21 (0.41)	0.17 (0.38)
B1	0.00 (0.04)	0.74 (0.44)	0.23 (0.42)	0.03 (0.17)	0.33 (0.47)	0.14 (0.35)	0.02 (0.14)	0.31 (0.46)	0.12 (0.32)
B2	0.01 (0.09)	0.72 (0.45)	0.24 (0.43)	0.03 (0.18)	0.23 (0.42)	0.13 (0.34)	0.09 (0.29)	0.35 (0.48)	0.13 (0.34)

Note. Schools A2 and B1 were basic-education schools while B2 was a boarding school. At B2, no Kinyarwanda lessons were observed. The numbers in brackets refer to standard errors associated with scale scores.

Within the present study, we gauged competence, utility, and cost (self-) evaluations through self-reports as well. Table 52 presents their two-level ICCs. The biggest part of the variances was also at the between-student level ranging from .542 (competence self-perceptions in mathematics) to .798 (utility perceptions in mathematics). Even though these levels appear to be more in line with the scope of between-student variances of self-reported engagement provided by the literature, they also tend to be higher than those ICCs that are reported for motivational constructs in the literature. There, ICCs reported were between .27 and .44 in the case of competence self-evaluations (e.g., Ahmed, Werf, et al., 2010; Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021), between .19 and .36 in the case of value beliefs (e.g., Dietrich et al., 2019; Salmela-Aro et al., 2021), and between .28 and .55 in case of cost perceptions (e.g., Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021). Hence, whereas motivation reported by Rwandan students was mainly driven by between-student differences, their engagement levels were mainly more driven by situational differences. Future research in Rwanda needs to operationalize engagement as a non-binary variable. In that case, the within-variance does not need to be proxied through the standard logistic distribution. It would be interesting to explore to what degree the high level of within-lesson variability would hold.

To what extent do classroom activities predict Rwandan students' behavioural engagement in class (hypothesis 2)?

As implied by table 51 on page 312, teacher explanations and individual seatwork constitute the most common classroom activities across schools and subject. Both accounted for around 27% of lesson time categorized. Comparing both in terms of scope shows that teacher-centred activities (42%) are around twice as common as student-centred tasks (21%) within lessons observed. Comparing these results with findings from other studies suggests that Rwandan lower-secondary classrooms are less teacher-centred and more student-focussed in their activities. Sakata (2022) found teacher-centred activities to account for 81% of the lesson time at Tanzanian primary schools whereas student-centred activities accounted only for about 14%. Observations at Ethiopian primary schools led Frost and Little (2014) to attribute 75% of lesson time to teacher-oriented activities and only about 15% to student-centred activities.

Table 52: Between/ within variance within engagement and motivation (manuscript five)

Scale	Between lessons (level 2)		Between students (level 3)	
	Estimate	%	Estimate	%
English				
Engagement	**0.22 [0.18; 0.30]**	.064	**0.01 [0.00; 0.04]**	.004
Competence	**2.30 [1.38; 4.16]**	.373	**3.87 [2.25; 7.19]**	.627
Utility	**1.01 [0.60; 1.74]**	.331	**2.04 [1.15; 3.72]**	.669
Costs	**1.85 [1.20; 2.99]**	.448	**2.29 [1.39; 3.88]**	.552
Kinyarwanda				
Engagement	**0.16 [0.10; 0.28]**	.045	**0.19 [0.08; 0.34]**	.051
Competence	**1.08 [0.55; 2.20]**	.267	**3.00 [1.29; 6.67]**	.733
Utility	**2.47 [1.08; 5.59]**	.351	**4.54 [1.81; 10.93]**	.649
Costs	**1.02 [0.50; 2.25]**	.227	**3.52 [1.71; 7.41]**	.773
Mathematics				
Engagement	**0.20 [0.17; 0.25]**	.056	**0.08 [0.05; 0.11]**	.022
Competence	**2.86 [1.84; 4.78]**	.458	**3.53 [2.07; 5.94]**	.542
Utility	**2.08 [1.33; 3.37]**	.202	**8.53 [4.85; 14.54]**	.798
Costs	**1.49 [1.01; 2.30]**	.427	**2.53 [1.30; 3.18]**	.573

Note. ‘Comp’ = competence states; ‘Daily’ = daily-utility states; ‘Cost’= cost states. ‘ICC’ = intracluster correlation. Statistical Analysis: multilevel CFA. Estimator: MCMC. Model fits: English: Sample size = 702; PPP = 0.422, *ƒ* (95% CI) = -36.42 - 44.42; Kinyarwanda: Sample size = 374; PPP = 0.436, *ƒ* (95% CI) = -35.76 - 45.26; Mathematics: Sample size = 773; PPP = 0.180, *ƒ* (95% CI) = -24.84 - 66.49. Coefficients in bold are significantly different from zero.

Table 53 on page 316 presents the results for the multilevel conditional mean models (CMM’s) linking up engagement and classroom activities at level 1. All three models exhibited excellent model fit as indicated by the last two rows. Not all effects are insignificant but generally tend to be small. Effects are significant when the square brackets associated with a coefficient does not contain the zero value (those cases are highlighted highlighted). Given that we performed logistic regressions, the effects within table 53 are presented as log odds, which can be translated into odd ratios, and using the school subject-specific average engagement levels across schools as base rate, into probabilities (e.g., Sommet & Morselli, 2017). For example, the log-odds of the effect of group work on engagement in English is .06 [.01; .10] translates to an odds ratio of roughly 1.062. Given an average engagement rate in English across schools of 73% (SD = .45), the engagement rate based on this odds ratio is 74%. This shows that effects observed might have partly been significantly different from zero but in case very marginal. Across subjects, the only predictor that consistently significantly from zero, and thus the biggest predictor oof engagement, was group work with effect size from .06 [.01; .10] in English to .15 [.10; .20] in mathematics. The latter effect was also the biggest one observed across activities and subjects. A student partaking in mathematics-related group work would experience a probability of being engaged of 77% instead of 74%. Comparing the activity types across subjects, yielded mixed results. On average activities we associated with student-centred classroom practices (i.e., whole-class discussions and individual seatwork) exerted positive effects. However, the effects of whole-class discussions were only significant within mathematics but not within English and Kinyarwanda. Nonetheless, we considered hypothesis 2a as largely confirmed by the data. Teacher-dominant activities (i.e., copying notes and individual seatwork) predominantly exerted effects not different from zero, which we considered to be in line with hypothesis 2b. Only in English, we found individual seatwork to positively predict engagement. Individual seatwork increased probabilities to be engaged from 73% to around 75%. Williams (2019), for example, observed that especially basic-education students in Rwanda have only little command over the English language. It may imply they feel more comfortable to engage with those learning activities that does not require from them to talk (or listen) such as performing grammar tasks.

Table 53: Estimated coefficients of the structural models (manuscript five)

Predictors	English (n= 139)		Kinyarwanda (n= 68)		Mathematics (n= 162)	
	CMM	ML SEM	CMM	ML SEM	CMM	ML SEM
Level 1 (within lessons): classroom activities						
Round	**.04 [.00; .08]**	**.04 [.00; .08]**	.04 [-.01; .04]	.05 [-.03; .13]	.01 [-.03; .04]	.01 [-.03; .04]
Discussions	.01 [-.04; .06]	.02 [-.04; .07]	.06 [-.13; .44]	.07 [-.05; .17]	**.08 [.04; .12]**	**.08 [.04; .12]**
Group work	**.06 [.01; .10]**	**.06 [.01; .11]**	**.10 [.01; .65]**	**.10 [.01; .19]**	**.15 [.10; .20]**	**.15 [.10; .19]**
Copying notes	.01 [-.04; .07]	.02 [-.04; .07]	.06 [-.15; .45]	.07 [-.05; .18]	-.02 [-.06; .03]	-.02 [-.06; .03]
Seatwork	**.08 [.02; .13]**	**.08 [.02; .14]**	.08 [-.05; .51]	.09 [-.03; .20]	.04 [-.01; .10]	.04 [-.01; .10]
Explanations	.00 [-.06; .06]	.00 [-.06; .06]	.00 [-.28; .65]	.01 [-.12; .13]	**.09 [.04; .14]**	**.09 [.04; .15]**
Level 2 (between lessons) covariates						
Tuesday	-.19 [-.41; .08]	-.15 [-.42; .14]	.20 [-.02; .43]	.20 [-.03; .42]	.03 [-.14; .19]	.02 [-.14; .18]
Wednesday	-.08 [-.29; .15]	-.05 [-.29; .20]	.15 [-.10; .42]	.18 [-.08; .42]	.07 [-.08; .22]	.06 [-.07; .21]
Thursday	-.24 [-.51; .05]	-.19 [-.49; .14]	.17 [-.11; .47]	.20 [-.09; .50]	**.16 [.00; .33]**	**.16 [-.01; .32]**
Friday	-.08 [-.32; .19]	-.05 [-.34; .25]	**.50 [.22; .77]**	**.53 [.26; .78]**	-.02 [.18; .15]	-.02 [-.17; .14]
Late morning block	-.10 [-.26; .07]	-.08 [-.24; .09]	.22 [-.05; .48]	.25 [-.03; .50]	**.21 [.10; .32]**	**.21 [.09; .32]**
Afternoon block	.06 [-.11; .22]	.06 [-.10; .23]	.11 [-.19; .38]	.17 [-.12; .42]	.06 [-.06; .18]	.06 [-.05; .19]
Competence		.17 [-.17; .54]		-.18 [-1.45; 1.01]		.04 [-.16; .24]
Utility		.07 [-.25; .35]		-.02 [-.91; .90]		.02 [-.16; .20]
Cost		-.00 [-.23; .21]		-.03 [-.77; .86]		-.06 [-.27; .14]
Level 3 (between students) covariates						
School B1	**.42 [.02; .81]**	**.40 [.02; .79]**	**-.45 [-.67;-.13]**	**-.41 [-.67; -.09]**	-.05 [-.31; .20]	-.06 [-.30; .20]
School A2	**.57 [.13; .95]**	**.58 [.13; .96]**	NA	NA	**-.35 [-.59;-.11]**	**-.36 [-.60; -.10]**
School level	.17 [-.25; .62]	.19 [-.27; .63]	-.22 [-.51; .10]	-.24 [.53; .09]	.03 [-.21; .27]	.04 [-.21; .27]
Gender	.11 [-.23; .47]	.11 [-.23; .46]	.20 [-.07; .46]	.22 [-.08; .47]	**.27 [.06; .46]**	**.26 [-.05; .46]**
PPP	.578	.04	.500	.07	.451	.057
ƒ (95% CI)	-21.56-17.69	-5.44–121.50	-18.52-22.96	-18.04-117.33	-18.93-18.78	-16.41-121.02

Note. 'CMM' refers to the conditional mean models estimated to answer research question 2. 'ML SEM' refers to multilevel structural equation models estimated to answer research question 3. Credibility intervals are in parentheses. Bold coefficients are significantly different from zero)

As per hypothesis 2c, we expected teacher giving explanations to the class to predict students' probabilities of being engaged. As table 53 shows, the data does not support this assumption. Only in mathematics, teacher explanations were a significant predictor. Probabilities of being engaged increased to around 76% on average. Hence, we found the three hypotheses related to classroom activities supported by the data only in mathematics but not in English, and Kinyarwanda. Within mathematics, our results replicated findings reported by others (Olivier et al., 2020; Skinner & Belmont, 1993). How to interpret the small effect sizes associated with classroom activities? At face value they may come across as negligible. However, engagement states may not be independent of each (e.g., Vasalampi et al., 2020). Across time, engagement is best understood as a series of states (i.e., process) where on-task behaviour at point t may influence on-task behaviour at point $t + 1$ (e.g., Schmitz, 2006). Hence, any effects on engagement at any point in time may thus unfold multiplier effects that sustain into the future for some periods. To the extent that conducive contextual factors, amongst others in form of supportive classroom practices, sustain, the effects may even accumulate over time, which would only be of benefit to students. Future longitudinal research in Rwanda, and elsewhere, need to further investigate this using autoregressive models (e.g., McNeish & Hamaker, 2020). Future research should also focus not only on whether or not a type of classroom activity was given but also on quality this type of classroom activity was implemented in class.

To what extent is Rwandan students' behavioural engagement in class related to their lesson-specific motivational self-reports (hypothesis 3)?

Table 53 also presents the coefficients with the structural equation models we estimated (see columns 'ML SEM'). They represent extensions of the conditional models in that we included the latent factors on competence, utility, and cost (self-) evaluations as level-2 predictors of engagement. As indicated by the last two rows in each column, model fit deteriorated significantly. Even though the posterior predictive p-value (PPP) dropped below .05 in case of English, we still considered all model fit acceptable given that credibility intervals presented in the last row contained the zero value. Within none of the models any of the self-reported motivational measures turned out to be a significant predictor of engagement. This was not in line with hypothesis 3 and with findings reported in the literature (Olivier et al., 2020). We did not consider this as evidence that the theory (e.g., Eccles & Wigfield, 2020) predicting task behaviour through competence, value, and cost factors does not hold within the Rwandan context. We rather assumed that the way engagement and the motivational variables resulted in some form of mismatch. The fact that variances associated motivational self-reports were mainly due to differences between students whereas variances associated with engagement appeared to originate especially at the within-lesson level (see table 52) may be the most obvious indication of this mismatch. One mismatch occurred in terms of the level of specificity (Swann et al., 2007). We tried to address it through parameterizing a multilevel structural equation model with engagement as a level-1 outcome and motivational self-reports as level-2 (i.e., with a random-intercept parameterization; Heck & Thomas, 2020). However, other differences remained such as the one within time orientation. Whereas engagement was gauged momentarily during lessons, competence, utility, and costs (self-) evaluations were collected after lessons. Through collecting students' motivational self-reports immediately after a lesson ended, we hoped to be able to proxy the typical motivational states students experienced that lesson. However, this assumption may not have held. To arrive at those lesson-specific typical motivational (self-) evaluations students would have to aggregate all individual states into one rating, something very challenging cognitively. Instead, students may have employed mental "short cuts” such as recalling just the last salient moment of a lesson to arrive at one rating. Even though researchers employed lesson-specific self-reports as measures of motivational states, it appears to be hardly studied to what extent students are actually good at gauging their lesson-specific typical motivational states retrospectively through one global rating. Our results let us believe that in the case of Rwanda, and potentially elsewhere, there is some discrepancy. Future research should also explore to what degree external ratings of students’ engagement levels does match their self-reported engagement levels. Some studies have shown that they may not correlate as highly as they ideally would (e.g., K. A. Fuller et al., 2018).

Table 53 on page 316 also presents the log-odds of the covariates we included into the different models. Time did not emerge as a relevant predictor of engagement levels besides some few exceptions. On some days, the probability of being engaged was significantly higher in Kinyarwanda (Friday) and in mathematics (Thursday). In mathematics, engagement also appeared to be slightly higher in the first hours of the day. Interestingly, school levels were not a predictor at all. In two respects, our findings seem to depart from what is discussed especially within the Western literature. First, in contrast to other studies we did not find gender to be an important predictor of engagement in Rwanda. Moreover, where gender mattered (mathematics) effects were not negative (e.g., Ansong et al., 2017) but positive in the sense that girls exhibited higher engagement on average. Second, evidence suggests that within many other parts of the world students exhibit decline within engagement already between grade seven and nine (e.g., Lam et al., 2016). Our data suggest otherwise for the case of Rwanda. School level was not a significant predictor within any of the three subjects. The only important covariate appeared to be school type. Basic-education schools were associated with higher engagement probabilities in English but lower probabilities in Kinyarwanda and mathematics than the boarding school. This is in line with

Bulla et al. (2022b) who found structural differences within motivational relations between basic-education and boarding school students.

Conclusions

Within the present study, we combined self-reports on lesson-specific competence, utility, and cost (self-) perceptions provided by students after class with in-situ observations of both classroom activities these students encountered as well as engagement states they exhibited during class. We found students to generally exhibit high levels of behavioural engagement while states of verbal engagement in particular were very uncommon. Especially in mathematics, we found activities we associated with student-centred practices as well as teacher explanations to exert small but significant effects on engagement while those activities we considered teacher-dominant did not show any significant effects. Those findings indeed suggest that student-centred pedagogies may constitute a positive influence on students' learning as has oftentimes been argued within the international-development discourse on education. By contrast, we did not find any effects of students' lesson-specific self-reports on engagement. We take this as indication that methodological differences in the way engagement and its potential motivational antecedents have been operationalised may explain this finding. Future research needs to further explore links between motivation and engagement as its postulated outward manifestations amongst Rwandan students.

6. General discussion

In the present DPhil project, a first attempt was made to empirically explore the internal as well as external nomological net of achievement motivation within the Rwandan school context. The nomological net is a schematised description of both the internal structure and external relations of a concept of interest (Byrne, 1984; Cronbach & Meehl, 1955). Rwanda is a country within Africa, a continent that has been generally underrepresented within mainstream motivational research (e.g., Cornelius-White, 2007; Kriegbaum et al., 2018; Lei et al., 2018; Möller et al., 2009, 2020; Orth et al., 2020; Roorda et al., 2011). It was for this reason that the present DPhil project was carried out within an African country. Through carrying out motivational research in Africa, the present DPhil project contributed to further diversifying the sample base upon which current models and assumptions concerning achievement motivation have been tested. As a psychological construct, achievement motivation is culturally dependent. Cultural dependency implies that the way achievement motivation manifests itself may vary across cultures. Theory building must take into account this cultural dependency. Only when theories have been formed and tested based on a sample base that is reflective of the cultural diversity of the human race, can their claims be expected to apply across the globe. Through examining key aspects of the internal as well as external nomological net of achievement motivation within the Rwandan school context, this DPhil project thus tested current conceptualisations of achievement motivation within a ‘novel’ research context. In addition to this theory-related rationale for carrying out the DPhil research Africa, there was a policy-related rationale as well. In general, motivational interventions have been found to have the potential to be effective ways to improve educational outcomes including academic achievement. They also have been found to be cost-effective interventions (e.g., Lazowski & Hulleman, 2016; Yeager &

Walton, 2011). Against this background, it is surprising that motivational interventions do not form part of the toolbox of educational programming to address the learning crises in many parts of the developing world inside and outside Africa. One explanation could be the dearth of empirical evidence on the links between achievement motivation and in particular academic achievement. The research focus within this DPhil project was selected based on the hypothesis that academic achievement explains academic achievement amongst students. Through examining these links within the African context, the present DPhil could have contributed towards building the case for shifting the policy focus towards motivational interventions as a means to improve learning outcomes in low-income countries. Finally, through studying motivational processes within Rwandan classrooms, it was also intended to obtain first insights into what aspects of achievement motivation may constitute promising entry points for intervention research around achievement motivation. Thus, based on theory- and policy-related grounds, Africa was chosen as a research context for this DPhil project. Within Africa, Rwanda was selected because from an educational point of view it has been a well performing but not atypical low-income country within Africa. It has been performing well in terms of attaining universal enrolment amongst children while facing the challenge of poor academic performance of its schooling system, which is typical for many low-income countries. Hence, the examination of the relations between achievement motivation and academic achievement as well as behavioural engagement was not only of theoretical but also of practical relevance interest.

The results of the examinations of the internal and external elements of the nomological net are summarised in the form of five manuscripts targeted at academic journals. They are presented in chapter five. The purpose of this chapter is to provide a general discussion of the patterns that emerge across the five manuscripts. Given that the nomological net within and around achievement motivation was used to organise the research objectives pursued within the context of this DPhil project, it also constitutes an adequate way to structure the general discussion of results in this chapter. The chapter's first two sub-sections focus on the internal aspects of the nomological net. First, the horizontal structure of achievement motivation is discussed. Reflecting on both the qualitative and the quantitative results, the level of multidimensionality of achievement motivation within the Rwandan context is compared with its current conceptualisations within the mainstream literature. Second, the multilevel structure of achievement motivation within the Rwandan context is discussed. The key question in this respect is to what extent motivational states exhibit the same level of intrapersonal variability observed in other parts of the world. Sections 6.3. to 6.5. are about the external elements of the nomological net around achievement motivation observed in Rwanda. The third sub-section discusses the extent to which achievement motivation is predicted by external variables, i.e., prior school grades as well as classroom activities. Within manuscript two and three, prior school grades were related to different ways of operationalising school subject-specific motivation (i.e., subject-specific self-reports versus average motivational states). The comparison of both operationalisations is the subject of sub-section 6.4. Sub-section 6.5. collates the evidence on the potential "meaning" (Benson, 1998, p. 11; Crocker & Algina, 2006) of achievement motivation within the Rwandan context. Meaning is discussed with regards to behavioural engagement as well as academic achievement. As for the latter, data on the predictability of academic achievement through achievement motivation that was collected as part of the qualitative-exploratory first study is discussed. The final sub-section 6.6 draws the overall conclusion of this DPhil project. Throughout sub-sections 6.1. to 6.6., research limitations and future research needs are identified. Thus, limitations are not discussed through a separate sub-section.

6.1. Motivation in Rwanda: multi-dimensional but less nuanced

Within the literature, achievement motivation is generally conceptualised multi-dimensionally (e.g., Schunk et al., 2013). Within the tradition of the expectancy-value theory, the model of achievement motivation relied upon in the context of this DPhil project, this multidimensionality is represented through students' success expectations, value perceptions, and increasingly cost evaluations (e.g., Eccles et al., 1983; Eccles & Wigfield, 2020). Moreover, the main categories of value and cost evaluations were further divided into different sub-dimensions (Brownlee et al., 2004; Flake et al., 2015; Gaspard et al., 2015), which made the expectancy-value model of achievement motivation a highly nuanced framework of students' cognitive-affective experiences. Within the qualitative-exploratory study reported in manuscript one, this multifaceted model of achievement motivation was applied to the descriptions of motivational (and confidence-related) experiences provided by Rwandan secondary students. The results suggest that Rwandan students' cognitive-affective experiences are indeed multidimensional as well. By and large, most student descriptions could be categorised as success expectations, value perceptions, or cost (self-) evaluations.[30] Thus, at the level of key categories, Eccles et al.'s model was also replicated within the Rwandan school context. However, two caveats are to be highlighted. First, the research design within the qualitative-exploratory study reported in manuscript one was directly derived from Eccles et al.'s expectancy-value theory. The questions used to probe students to describe their cognitive-affective experiences were designed in such a way as to mirror the expectancy, value, and cost elements of Eccles et al.'s model. The coding scheme employed

[30] In addition to the three motivational dimensions of success expectations, value perceptions, or cost (self-)evaluations, personal goals were identified as well. This suggests that the expectancy-value theory is not the only model to describe achievement motivation within the Rwandan context. Future research should focus on the nomological net within and around achievement motivation conceptualised through achievement goals (e.g., Elliot & Hulleman, 2017).

to code the motivation data in particular was directly derived from the highly nuanced conceptualisations of expectancy, value, and cost (self-)evaluations put forward in the recent literature. Thus, different (qualitative) research designs may reveal different conceptualisations of Rwandan students' experiences of achievement motivation.

Second, even though Eccles et al.'s model was replicated at the level of the three main categories, Rwandan students' cognitive-affective experiences appeared to be less nuanced at the level of the sub-categories. This was observed in the case of cost as well as value perceptions. Within the literature, value beliefs are increasingly represented as an eight-dimensional construct (e.g., Gaspard et al., 2015) whereas cost is conceptualized through three (e.g., Gaspard et al., 2015) to four (Flake et al., 2015) dimensions. Figure 18 on the next page represents the subject-specific as well as lesson-specific factor solutions of achievement motivation that emerged in the context of this DPhil project. The former consists of three factors (i.e., expectancy, daily utility, and cost beliefs) whereas the latter comprises of four factors (i.e., expectancy, daily and school utility as well as cost evaluations). Initially, a facet that was labelled perceived uselessness emerged as a factor as well. As a concept, it may also be related to what is discussed as "useless schooling" (Ailei, 2019) within ethnographic educational research. It describes the notion that "studying is not useful, and it is better not to study" (p. 186). As it has been argued in manuscript one, perceived uselessness could be interpreted as the perceived lack of attainment value, which represents the "personal importance" (Eccles, 2005, p. 109) attached to scholastic learning. During follow-up examinations (see section 5.2), perceived useless appeared to be untenable as a motivational factor. Within the corresponding exploratory factor studies, perceived useless loaded on items across different school subjects rather than on self-report items only within a specific school subject. Unlike other motivational factors such as expectancy beliefs/ evaluations, perceived useless was not subject specific. It thus resembled a method factor. In fact, items used to operationalize perceived useless used negations, which might have caused confusions amongst secondary students filling the self-report questionnaires. Hence, perceived uselessness, as operationalized within this DPhil project, was not further pursued during the main study solely on methodological rather than conceptual grounds. This in turn also implies the possibility that attainment value may indeed of relevance within the Rwandan context. It may just manifest itself in form of the complete absence of attainment value. Future research needs to further explore this.

Figure 18: The final motivational factors operationalized through the self-report scales

As the example of perceived uselessness highlights, cost and value beliefs as well as evaluations within the Rwandan school contexts may be more nuanced than the empirical findings reported in this DPhil project may suggest. One explanation for the smaller extent of multidimensionality of cost and value beliefs as well as evaluations within the context of this DPhil project might have been the size of the students sampled employed during the exploratory factor studies. The initial exploratory factor-analytical studies (EFA I and EFA

III) were rather small in scope with $n_{EFA\ I} = 371$ and $n_{EFA\ III} = 197$. In particular given that data were ordinal in nature in both cases, much larger sample sizes would have been in order (e.g., Fabrigar & Wegener, 2012; Watkins, 2018). Thus, replication of the quantitative exploratory factor work with larger samples might result in higher-dimensional conceptualisation of achievement motivation exhibited by Rwandan students. Such studies may also be able to distinguish between intrinsic value as well as expectancy beliefs empirically, something that did not emerge within the examinations of this DPhil project. Also, within higher powered studies one may find daily and general utility as distinct motivational factors (e.g., Gaspard et al., 2015) as well as cost to be a more nuanced driver of motivation (e.g., Flake et al., 2015). Within this DPhil project, items of daily and general utility loaded just on one factor, which was eventually called daily utility (see section 5.2). Similarly, even though operationalized multidimensionally, cost emerged as a single factor only.

Nonetheless, there are theoretical reasons to assume that cost and in particular value beliefs as well as evaluations may not be as nuanced amongst Rwandan lower secondary students as they have been found to be amongst Western students. The observed lack of relevance of job utility may not be entirely counterintuitive. All study participants attended lower secondary classes and thus may have expected to attend school for at least another 3 years. They may have therefore not deemed it necessary to ponder upon the usefulness of any subjects, let alone activities within subjects, for one future life. Another explanation could be that the survey questions were explicitly about the reasons for being (de-) motivated and (un-) confident about school exercises. Even though students were asked to name those subjects they generally feel (de-) motivated and (un-) confident about, this wording of the survey questions may have nudged students to think along those lines that did not let them truly reflect all potential benefits of education. For example, in school there may generally be not many exercises that, in isolation, are particularly useful for one's own future professional career. Thus, explicit questions about the usefulness of a subject may have yielded more answers related to daily, general, and job utility.

There is additional, implicit, evidence that the manifestation of achievement motivation may not greatly differ from manifestations of achievement motivation from other parts of the world. Especially within self-concept research, there has been an ongoing discussion about the extent to which competence-related self-beliefs consists not only of a cognitive but also of an affective dimension (Arens & Hasselhorn, 2015; e.g. Marsh, 1999; Marsh et al., 2002; Pinxten et al., 2014). This issue also appeared within the quantitative data used within the EFA I study. Across the different factor solutions, the self-report items used to gauge students' intrinsic value perceptions loaded on both the value as well as expectancy factor. Future research, amongst others with bigger samples, needs to address to what extent intrinsic value can be teased out into a separate factor. Based on the qualitative and quantitative data analysed within the context of the present DPhil project, however, achievement motivation exhibited by Rwandan secondary students appeared to be multidimensional but less nuanced especially when compared with conceptualisations currently discussed within the expectancy-value theoretical literature. It appeared to be multidimensional because it, too, consisted of expectancy, value, and cost factors (compare figure 18 with figure 1). It appeared to be less nuanced because the different facets of value and cost evaluations were less common within the qualitative data (see manuscript one) and did not emerge as distinct factors within the quantitative data at all (see figure 7).

6.2. Motivation in Rwanda: multileveled but less intrapersonally variable

The qualitative data analysed in manuscript one suggests that Rwandan students form cognitive-affective perceptions around schooling at different levels of specificity. Within their descriptions of motivational and confidence experiences, some students referred to "tasks" or "exercises" while others referred to subjects such as mathematics. It is anecdotal evidence in support of a multileveled manifestation of achievement motivation within Rwanda. Within the subsequent scale-development process, scales for both the school subject-specific and for the lesson-specific level were developed. They in turn allowed studying achievement motivation at the different levels of specificity. In line with the motivational literature, motivational states were assumed to be intrapersonally variable in the sense that they were assumed to vary across lessons. To examine the level of intrapersonal variability within motivational states, the intracluster correlations (ICC; e.g., Heck & Thomas, 2020; Raudenbush & Bryk, 2002) were estimated. They are reported within manuscripts three to five. The studies reported within the manuscripts varied in terms of sample size (n_3 = 446 students, n_4 = 868 students, and n_5 = 188 students). The studies also varied in terms of the constructs included. Whereas manuscript three only focused on competence and cost (self-)evaluations, manuscripts four and five also included daily-utility perceptions. None of the studies included school utility since models estimated in manuscripts three to five generally did not converge when school utility was included. For the same reason, daily utility was not included in the models reported in manuscript three. They generally did not converge in the case of daily utility. For comparison purposes, the data set reported in manuscript four was used to estimate the ICCs for competence, daily utility, and cost (self-)evaluations as well as school utility perceptions within English, Kinyarwanda, and mathematics. The table 3 on page 95 highlights that ICCs across constructs and subjects were substantial, but not exorbitantly high, implying that around 53 to 73 per cent of the variances originated at the between-lesson level. These results are replicated based on the samples analysed within manuscripts three to five. Thus, the level of intrapersonal variability within motivational states observed amongst Rwandan students is substantial (i.e., between 27 and 47 per cent) but appears to be lower than what is usually reported for Western samples. As discussed in the literature review as well as manuscript four, the share of variance attributable to the lesson-specific level tends to be between 56 and 73 per cent in the case of competence self-evaluations (e.g., Ahmed, Werf, et al., 2010; Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021), between 64 and 81 per cent in the case of value beliefs (e.g., Dietrich et al., 2019; Salmela-Aro et al., 2021), and between 45 and 55 per cent in the case of cost perceptions (e.g., Dietrich et al., 2017, 2019; Malmberg et al., 2013; Parrisius et al., 2021; Salmela-Aro et al., 2021). The differences are particularly stark in case of daily utility. The between-lesson variances accounted for around just 25 per cent in the Rwandan case, which is far smaller than the 81 per cent reported by Dietrich et al. (2019) based on a German sample, and 64 per cent reported by Salmela-Aro et al. (2021) based on a Finish sample. This may suggest that especially Rwandan daily-value perceptions are less susceptible to lesson-specific influences. Once Rwandan students have formed utility perceptions, they tend to be rather stable. One explanation for this could be rooted within general attitudes towards education amongst Rwandans. In Rwanda, there is a strong perception that schooling is a key avenue towards socio-economic betterments. This perception is not centred around schooling as the place that imparts knowledge and skills. It is rather centred around schooling as the place where students obtain school degrees, which in turn may lead to a better life (Honeyman, 2015b). Hence, perceived utility may be determined especially in reference to the perceived value of the school degree one can expect to eventually obtain. This in turn may explain why situational factors may be less important in shaping one’s utility perception. Future research needs to examine to what degree this explanation indeed holds true in the Rwandan context.

Two additional comments on the level of temporal stability are in order. The first one concerns behavioural engagement. Whereas motivational states at the lesson-specific level exhibited only moderate levels of intrapersonal variability, the estimated variability within behavioural engagement was extremely high. The estimation procedure outlined in manuscript five suggested that only around 90 per cent of the variance within engagement originated at the within-lesson level. However, operationalising behavioural engagement through an on-task/off-task coding scheme required proxying the level-1 variance through the variance of the standard logistic distribution. However, using ${\pi^2}/{3}$ as a proxy for the level-1 variance within three-level logistic models might have underestimated the level-2 and level-3 ICC. Thus, operationalisations of engagement through more nuanced coding schemes may allow more robust ways of estimating level-1 variance, which in turn may yield different levels of intrapersonal variability within behavioural engagement. Future research needs to explore this question further. The second comment on temporal stability concerns subject-specific motivational self-reports. Within the literature, motivational beliefs are reported to be rather stable within (e.g. Marsh et al., 2005; study 1) as well as across school years (Lauermann et al., 2017; Marsh & Martin, 2011; Pinxten et al., 2014). Within the context of the present DPhil project, this question was not addressed empirically. Given that motivational states reported by Rwandan students are less susceptible to short-term fluctuation, their motivational beliefs may exhibit slightly different temporal patterns as well. Future research should therefore also explore to what degree school subject-specific motivational beliefs such as expectancy, value, and cost (self-)evaluations are stable across time within the Rwandan context. To sum up this section, based on the results gathered through this DPhil project, achievement motivation exhibited by Rwandan lower-secondary students appears to be not only a multidimensional but also a multilevel construct. Even though multileveled, motivational states appeared to be less shaped by situational factors. The latter expressed itself through ICCs that were higher than commonly reported for expectancy, value, and cost (self-)evaluations in the literature. Hence, even though achievement motivation was found to be less nuanced and less intrapersonally variable, its local manifestation in Rwanda matched its conceptualisation as a multidimensional and multileveled concept within the wider literature.

6.3. Motivation in Rwanda: externally predicted but predictability varies across motivational constructs and contexts

The external nomological network around achievement motivation consists of two sets of variables. First, it consists of variables that are assumed to predict achievement motivation, and, second, it consists of variables that are hypothesised to be predicted by achievement motivation (e.g., Byrne, 1984). The former is subject of manuscripts two to four and discussed in this as well as the next sub-section. Within the present DPhil project, achievement motivation was explained through prior school grades as well as through classroom activities. Relations with school grades were examined at the subject-specific level whereas relations with classroom activities were studied at the lesson-specific level. A key finding across the different manuscripts has been that relations were moderated by different factors. These factors included the type of motivational construct being examined the school type involved, and the school subject concerned. Prior school grades were found to predict achievement motivation especially in the form of school subject-specific self-reports (manuscript two) but partly also in the form of average motivational states (manuscript three). As for school subject-specific self-reports, the most fundamental logic of the GI/E model was replicated within the Rwandan case as well. Prior school grades predicted subsequent achievement motivation within school subjects (i.e., chemistry, English, Kinyarwanda, and mathematics) as well as across school domains (i.e., quantitative versus language subjects). However, GI/E model relations were moderated by school type. Whereas the model explained expectancy as well as cost beliefs in the case of boarding school students, it predicted cost beliefs only in the case of basic-education students. Boarding school students exhibited higher levels of school subject-specific self-reports especially in the case of expectancy (apart from Kinyarwanda) as well as cost beliefs (apart from chemistry). The sample used in manuscript two was somewhat balanced in terms of the numbers of students from both basic-education and boarding schools. The subsequent studies were much more imbalanced in this respect. In subsequent studies, the only boarding school left in the samples was institution B2, which was also the smallest of all schools studied with a total number of participants of 107 only. Hence, whereas the moderating effects of school type was examined using multigroup analysis in the study reported in manuscript two, all subsequent studies controlled for the effects of school type by including school-specific indicator variables into the structural models estimated. Only the Kinyarwanda-specific models were not controlled for the dummy effects of school B2 since no lesson-specific motivational data was collected at school B2. Similarly, to what was observed in the subject-specific motivational self-reports, students from basic-education schools tended to have lower average motivational states (manuscript three) and lower lesson-specific motivational states (e.g., manuscript four). As already alluded to throughout the different manuscripts, in Rwanda students that have performed well at national examinations are more likely to be accepted by boarding schools. Thus, on average, students at boarding schools tend to perform better compared to basic-education students in terms of school marks (see manuscript two). Moreover, students at boarding schools are expected to know that they attend schools that are considered to be for those that perform better academically. Both effects combined may explain why students at boarding schools tend to exhibit more favourable motivational beliefs and states. In the case of behavioural engagement, results were more mixed (manuscript five). In English, basic-education students exhibited higher levels and in mathematics they exhibited lower levels of behavioural engagement in class.

Relations concerning achievement motivation were also moderated by school subjects. For example, average competence and cost states were predicted by prior school grades within mathematics but not at all within Kinyarwanda (manuscript three). School subject turned out to be a significant moderator within the study reported in manuscript four as well. Whereas all types of classroom activities exhibited only positive effects on students' lesson-specific (self-)evaluations in English and in particular in mathematics, respective effects were mainly negative in Kinyarwanda. A similar pattern also emerged in the study reported in manuscript five. Classroom activities were observed to be predictors of behavioural engagement in mathematics and English. By contrast, behavioural engagement exhibited by students were largely unrelated to the different types of classroom activities within Kinyarwanda. Patterns concerning achievement motivation were also moderated by motivational construct. Within manuscript two, the GI/E model applied especially to expectancy (boarding school students) and cost beliefs (basic-education and boarding school students) but not to daily utility. This in turn provides additional support to the postulate discussed in section 6.1 that in Rwanda daily utility may be largely determined by wider attitudes around education than by lesson and school-subject-specific variables such as prior academic performance. The effects of classroom activities on achievement motivation reported in manuscript four appeared to be moderated by an interaction by school type and motivational construct. For example, whereas competence states were hardly predicted by classroom activities in English, they were observed to be influenced by classroom activities in mathematics. What's more, cost states were favourably influenced by all classroom activities in mathematics. By contrast, cost evaluations were hardly related to classroom activities in Kinyarwanda. Within manuscript five, the effects of expectancy, value, and cost (self-)evaluations on students' behavioural engagement was consistent in that none of them exerted any statistically significant influence.

Across the different studies, it was also controlled for age and gender. Gender was not always a statistically significant predictor of motivational beliefs. Female students tended to have lower expectancy beliefs, higher value perceptions, and more negative cost beliefs. Older students tend to have lower expectancy beliefs in Kinyarwanda (basic-education school) and chemistry (boarding schools), and more negative cost beliefs in chemistry (boarding schools), English (basic-education school), and Kinyarwanda (basic-education and boarding schools). In the case of average motivational states (manuscript four), gender was no predictor at all. As for age, older students tended to exhibit slightly lower average states only in Kinyarwanda. The finding that age and gender was not an important factor was also replicated in the study reported in manuscripts four and five. Only in mathematics did female students tend to exhibit slightly higher engagement levels compared to boys. The finding that females tend to exhibit slightly less favourable motivational beliefs (manuscript two) at the school subject-specific level while not exhibiting any differences with boys in terms of their motivational states (manuscript three and four) has been observed in the wider motivational literature (Götz et al., 2013). As already explained, school subject-specific self-reports are best understood as generalisations about oneself within the context of a given school subject whereas state self-reports may be more reflective of one's actual experience at a given point in time (e.g., Bieg et al., 2014; Götz et al., 2013; Roos et al., 2015). As for subject-specific self-reports, these generalisations about oneself may also be shaped by gender-based stereotypes (Bleeker & Jacobs, 2004; Spencer et al., 1999) students encounter in Rwanda and elsewhere (Bleeker & Jacobs, 2004; Frenzel et al., 2010; Rozek et al., 2015; Tiedemann, 2000). In America, for example, Jacobs and Eccles (1992) found that parents' competence evaluations of their children predicted the competence beliefs of these children. In this respect, they also found parents to differentially evaluate the competencies of boys and girls especially in math. These parental stereotypes were found to explain subsequent competence self-perceptions of boys and girls. The data collected as part of this DPhil project suggest a similar dynamic to be at play within the Rwanda context. To summarise this section, as the relations with predictor variables prior school grades as well as classroom activities show, achievement motivation in Rwanda is a “heterogenous” concept. It is heterogenous in the sense that the different external relations with school grades and classroom activities are moderated by the type of motivational construct (i.e., expectancy, value, and cost perceptions) focused on, and by the context (i.e., school type and subject) these relations are studied in. In part, student-level variables such as age and gender seem to matter, too. These dynamics need to be factored in especially when achievement motivation is used as a concept to guide educational programming. Thus, future research should aim at better understanding these dynamics.

6.4. School subject-specific motivation in Rwanda: self-reported beliefs and average states are related but distinct operationalisations

The representation of achievement motivation as a multidimensional and multileveled concept is rather common within the educational literature. As further discussed especially within chapter two, its multilevel structure implies that achievement motivation operates amongst others at the state, and the trait level, as well as the school subject-specific level in between. The representation of motivational constructs that way has been common (e.g., Götz et al., 2006; Shavelson et al., 1976; Vallerand, 1997). The way to operationalise constructs at the school subject-specific level has increasingly become subject to scientific enquiry. Especially within personal research, the operationalisations through direct self-report measures as well as through average states have been compared. Within education, respective questions have been addressed as well, though often sporadically. The review of manuscripts two and three reported in chapter five provides opportunity for a systematic review of both operationalisations within the Rwandan context. Within manuscript three, both operationalisations were directly related. Factor correlational work suggests that associations between average competence and cost states on the one side and expectancy and self-reported cost beliefs on the other side are substantial but not approaching any maximum. This in turn would suggest that both operationalisations are related but not identical (Raykov & Marcoulides, 2011). In addition to the level of convergent validity examined within manuscript three, the GI/E model has been applied to both operationalisations. To the extent that the model would have predicted school subject-specific self-reports as well as average motivational states in a similar fashion this could have been taken as evidence that both constructs measure the same phenomenon. However, GI/E model relations pan out rather differently within both cases. As explained in manuscript two, the most fundamental GI/E model logic applies to school subject-specific self-reports especially within the case of boarding-school students. As for average motivational states, not even the most fundamental model logic is replicated. Only within the subject of mathematics did prior school grades predict subsequent average competence and cost states. This supports the assumption that both constructs measure different phenomena. Furthermore, relations with prior school grades are also differentially moderated by gender across both constructs. As explained in the previous section, female students tended to exhibit slightly lower levels of school subject-specific motivational beliefs. As explained, gender-based stereotypes may shape girls' self-generalisations. This dynamic does not seem to be at play when discussing average motivational states. There tends to be no difference between male and female students in terms of the level of average states exhibited. To sum up this discussion, school subject-specific motivational beliefs and average motivational states may be related but distinct constructs, nonetheless. The key question in this respect is to what degree both are relevant. Within the wider literature, school subject-specific motivational beliefs have been found to predict important educational outcomes such as academic achievement (e.g., Hailikari et al., 2008; e.g., Kriegbaum et al., 2018; Lavrijsen et al., 2021; Steinmayr & Spinath, 2009), or career choices (e.g., Lauermann et al., 2017). Future research, both in Rwanda and elsewhere, needs to examine to what degree average motivational states predict important educational outcomes such as academic achievement or career choices as well. If, and only if, they do, average motivational states may not only be an interesting statistical artefact (i.e., the mean of a series of self-reported motivational states) but constructs that hold “meaning” (Benson, 1998, p. 11; Crocker & Algina, 2006).

6.5. Motivation in Rwanda: ‘meaningful’ but more research is needed

The relations between prior school grades on the one side and different motivational constructs on the other side reviewed in section 6.3 were correlational in nature. To some extent, these associations may thus be indicative of relations between achievement motivation and subsequent school grades. Given that prior school grades were found to predict school subject-specific self-reports and partly average motivational states, this would imply that achievement motivation may also predict subsequent school grades (e.g., Hailikari et al., 2008; e.g., Kriegbaum et al., 2018; Lavrijsen et al., 2021; Steinmayr & Spinath, 2009). This type of evidence could have been used to build a particularly convincing case for the relevance of achievement motivation within education programming. As it seems, the policy focus has been increasingly shifting towards improving learning outcomes rather than increasing enrolment rates within the contexts of the developing world (e.g., Pritchett, 2013). Due to the Covid outbreak in Rwanda in early March 2020, all schools were subject to closures throughout the country. This in turn prevented the systematic study between expectancy, value, and cost (self-)evaluations on the one hand and subsequent school grades on the other. However, within manuscript one, correlational data are reported which provided first evidence on the existing relations between motivational constructs and academic performance. Students' perceived associations of school subjects with motivation, confidence, or the lack thereof, partly predicted school grades they obtained at the end of that school year. Subjects that were associated with confidence and especially with motivation were also associated with higher school grades. This was in line with what expectancy-value theory predicted (e.g., Kriegbaum et al., 2018). This is remarkable given that the way the terminology used was rather a crude conceptualisation of students' expectancy and value (as well as cost) beliefs. Nonetheless, they predicted school grades. This in turn may point to the general relevance of cognitive-affective factors for explaining learning outcomes not only in the Western world (Kriegbaum et al., 2018; Lavrijsen et al., 2021) but also amongst Rwandan students. This result may encourage future research on the nexus between motivation and learning outcomes, ideally a more robust and systematic way. The potential "meaning" (Benson, 1998, p. 11; Crocker & Algina, 2006) of achievement motivation within the Rwandan context may not only be indicated by the correlational data reported within manuscript one. It may also be implied by the qualitative data reported within the study of manuscript one. Being asked about the reasons why students work hard, a broad range of cognitive-affective factors were highlighted by Rwandan respondents. For example, almost 72 per cent of respondents providing descriptions of reasons for being engaged gave at least one statement that was about expectancy beliefs. This in turn provides further evidence that achievement motivation has relevance for students and their attempts to be academically engaged with their learning. Unfortunately, the qualitative results on relations between motivational factors and engagement were not replicated by the quantitative study reported in manuscript five. As manuscript five highlights, students exhibited somewhat high average engagement levels with being on-task around 75 per cent of the time. Potentially, this can be explained by Rwandan culture, which has been described as collectivist in nature. Being collectivist may imply a cultural emphasis on student effort and thus competence beliefs may be a relatively strong predictor of engagement. In fact, Honeyman's (2015b) qualitative study on the rollout of a new course on entrepreneurship at Rwandan secondary schools observed that students generally exhibit a "high degree of discipline and effort"(p. 166). Unlike theory predicts, engagement was not observed to be the "outward manifestation of motivation" (Skinner et al., 2009, p. 224) within the context of this DPhil project. None of the three dimensions predicted behavioural engagement. As discussed further in manuscript five, this could have potentially been due to a number of "mismatches" in the way how both engagement and motivation were operationalised. Future research should address these mismatches to explore to what degree the relations between the motivational factors as well as behavioural engagement remain insignificant. A first step in this respect may indicate a better understanding of to what degree external ratings as well as self-reported ratings of behavioural engagement are indeed congruent (e.g., K. A. Fuller et al., 2018; Peterson et al., 1984). Within their study, Fuller et al. (2018) triangulated the measurement of engagement by combining external observations with self-reports. They considered this attempted triangulation as unsuccessful "primarily because student self-reported engagement showed minimal correlation with engagement reported by an observer" (p. 14). They partly explained this lack of associations by the finding that students "pretended to participate" in about 23% of those instances where external observers rated them as being on-task (also see Peterson et al., 1984). Within Rwandan classrooms, it may be particularly easy for students to pretend being engaged during those classroom activities that may hardly require any mental involvement such as copying notes from the blackboard. Studies that combine observed engagement with self-reported engagement can then also include self-reported items on cognitive (Fredricks et al., 2004; Zimmerman, 1990) and emotional (Fredricks et al., 2004; Skinner et al., 2008) engagement. As already discussed, behavioural engagement may be a necessary condition for both cognitive and emotional engagement. It may also be differently related to self-reported motivation than cognitive and emotional engagement. In this respect, future research should also explore that external achievement motivation is related to verbal engagement. Given the very incidences of verbal engagement, it was not possible to study these associations within the context of this DPhil project.

How should one assess the “meaning” (Benson, 1998, p. 11; Crocker & Algina, 2006) of achievement motivation in Rwanda based on the DPhil data collected? The complexities of the potential relations between externally observed behavioural engagement at the situational level as well as retrospectively self-reported motivational states at the lesson-specific level might not have been adequately addressed within the context of this DPhil project. Given these methodological caveats, one cannot take the data reported in manuscript five as evidence against the hypothesised motivation-engagement nexus within the Rwandan context. It may just indicate additional, potentially more adequately designed research is necessary to examine these associations. However, qualitative data suggests that Rwandan students themselves explain their level of engagement through cognitive-affective factors. Furthermore, correlational data provides first evidence that achievement motivation related constructs indeed predict academic achievement. Hence, these different pieces of evidence may provide some first indication of the potential meaning of achievement motivation for Rwandan students’ learning in class.

6.6. Concluding remarks

What is the overall conclusion to be drawn for this DPhil project? Rwanda as a research context was chosen on theory-related and policy-related grounds. The former implied a comparison of the manifestation of achievement motivation within the Rwandan context with its current conceptualisations discussed in the (Western-borne) literature. The latter implied building a case for the systematic study of achievement motivation within the Rwandan context by highlighting the concept's relations with important educational outcomes of behavioural engagement and subsequent prior academic achievement. Admittedly, the results of manuscript five in section 5.6 do not help to build a case for achievement motivation as a future research field and as a theoretical lens to guide educational interventions within the Rwandan context. After all, no evidence on the relations between achievement motivation and behavioural engagement, as operationalised within the present DPhil project, was found. However, results obtained through this DPhil project are promising in other respects. First of all, the correlational data reported in manuscript one provides direct evidence on the significant relations between students' experiences of motivation and confidence within subjects on the one side and subsequent school grades on the other side. Second, even though achievement motivation in Rwanda appears to be less nuanced as well as less intrapersonally variable, its internal structure resembles current conceptualisations discussed within the wider motivational literature. This in turn can be seen as indirect evidence for the potential relevance of achievement motivation within the Rwandan context. If the Rwanda-specific manifestation of achievement motivation resembles the overall internal structure of manifestations observed in other contexts and achievement motivation has been found to predict behavioural engagement and academic outcomes in other contexts, then there is a case to be made to assume that future research may also show that achievement motivation predicts important educational outcomes such as school grades within the Rwandan school context. Achievement motivation may indeed constitute a framework that eventually helps to design impactful policy interventions to tackle various educational outcomes such as learning attainment of Rwandan students.

Even though this DPhil project provided some reasons to believe that achievement motivation may be a meaningful concept for Rwanda's schooling system, this DPhil wok certainly highlighted the need for substantial future research around achievement motivation amongst Rwandan students. As the different manuscripts presented in chapter five highlight, achievement motivation in Rwanda is far from being a “homogenous” concept. It is heterogenous in the sense that its internal and external manifestations vary across motivational dimensions (e.g., utility perceptions versus expectancy and cost (self-) evaluations), school context (e.g., basic education versus boarding schools), and school subject (e.g., mathematics versus language domain). For example, utility perceptions seem to be less intrapersonally variable compared to expectancy and cost (self-)evaluations. The GI/E model replicates within the Rwandan context. However, it does so differently between basic education versus boarding schools. Classroom activities predict students’ motivational experiences in class, but they do so differently across the different school subjects. Future research is thus required to better understand the dynamics around the different aspects of achievement motivation. Only when these dynamics are better understood, may achievement motivation have the potential to guide future education programming. This, however, is not Rwanda-specific. It is a general requirement to make motivational interventions work. In general, impactful motivational interventions can only be designed based on “a precise understanding of students’ subjective experience in school” (Yeager & Walton, 2011, p. 283). The need for additional research is no surprise. After all, it may be beyond the scope of a single DPhil project to develop a fully-fledged model of the internal and external manifestation of achievement motivation for Rwanda or any other country context. What are then the main takeaway points of this DPhil project? A fair summary may be that through this DPhil project, first direct and in particular indirect evidence was collated that suggests achievement motivation could be a concept that indeed holds “meaning" (Benson, 1998, p. 11; Crocker & Algina, 2006) for Rwanda’s school system. However, the way how to incorporate achievement motivation into Rwanda's school system in a meaningful way still requires future work. With this conclusion the present DPhil is drawn to an end.

Bibliography

Abadzi, H. (2007). *Absenteeism and Beyond: Instructional Time Loss and Consequences*. http://econ.worldbank.org.

Abbott, P., Sapsford, R., & Rwirahira, J. (2015). Rwanda’s potential to achieve the millennium development goals for education. *International Journal of Educational Development*, *40*, 117–125. https://doi.org/10.1016/J.IJEDUDEV.2014.12.007.

Adekunle, J. (2007). *Culture and customs of Rwanda*. Greenwood Press.

Adelman, M., Baron, J. D., Blimpo, M. P., Evans, D. K., Simbou, A., & Yarrow, N. (2015). *Why Do Students Learn so Little? Seeking Answers Inside Haiti’s Classrooms* [Document]. World Bank, Washington, DC.

AERA, APA, and N. (2014). *Standards for educational and psychological testing*.

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010). The role of competence and value beliefs in students’ daily emotional experiences: A multilevel test of a transactional model. *Learning and Individual Differences*, *20*(5), 507–511. https://doi.org/10.1016/J.LINDIF.2010.03.005

Ahmed, W., Werf, G., Minnaert, A., & Kuyper, H. (2010). Students’ daily emotions in the classroom: Intra-individual variability and appraisal correlates. *British Journal of Educational Psychology*, *80*(4), 583–597. https://doi.org/10.1348/000709910X498544

Ailei, X. (2019). “Useless Schooling” or “Hopeless Schooling”: An Ethnographic Study of Lower-Class Rural Parents’ Perceptions on the Value of Schooling [Article]. *Chinese Education and Society*, *52*(3–4), 186–207. https://doi.org/10.1080/10611932.2019.1667688

Alberto, P., & Troutman, A. C. (2013). *Applied behavior analysis for teachers* (9th ed.). Merrill ; Pearson Education [distributor].

Altmann, J. (1974). Observational Study of Behavior: Sampling Methods [Article]. *Behaviour*, *49*(3–4), 227–266. https://doi.org/10.1163/156853974X00534

Anderman, E. M. (2020). Achievement motivation theory: Balancing precision and utility [Article]. *Contemporary Educational Psychology*, *61*(1), 1–7. https://doi.org/10.1016/j.cedpsych.2020.101864

Ansong, D., Eisensmith, S. R., Okumu, M., & Chowa, G. A. (2019). The importance of self-efficacy and educational aspirations for academic achievement in resource-limited countries: Evidence from Ghana. *Journal of Adolescence*, *70*, 13–23. https://doi.org/10.1016/j.adolescence.2018.11.003

Ansong, D., Okumu, M., Bowen, G. L., Walker, A. M., & Eisensmith, S. R. (2017). The role of parent, classmate, and teacher support in student engagement: Evidence from Ghana. *International Journal of Educational Development*, *54*, 51–58. https://doi.org/10.1016/j.ijedudev.2017.03.010

Ansong, D., Okumu, M., Hamilton, E. R., Chowa, G. A., & Eisensmith, S. R. (2018). Perceived family economic hardship and student engagement among junior high schoolers in Ghana. *Children and Youth Services Review*, *85*, 9–18. https://doi.org/10.1016/j.childyouth.2017.11.008.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of*

School Psychology, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety. Within- and cross-domain relations with achievement and the mediating role of academic self-concept [Article]. *Contemporary Educational Psychology*, *51*(51), 240–252. https://doi.org/10.1016/j.cedpsych.2017.08.005

Arens, A. K., & Hasselhorn, M. (2015). Differentiation of competence and affect self-perceptions in elementary school students: extending empirical evidence. *European Journal of Psychology of Education*, *30*(4), 405–419. https://doi.org/10.1007/s10212-015-0247-8

Arens, A. K., Helm, F., Wolff, F., & Möller, J. (2020). Social and dimensional comparisons in the formation of German students' language self-concepts [Article]. *The Journal of Experimental Education*, *88*(2), 245–264. https://doi.org/10.1080/00220973.2019.1635069

Arens, A. K., & Preckel, F. (2018). Testing the internal/external frame of reference model with elementary school children: Extension to physical ability and intrinsic value. *Contemporary Educational Psychology*, *54*, 199–211. https://doi.org/10.1016/j.cedpsych.2018.06.003

Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2013). A short German version of the Self Description Questionnaire I: theoretical and empirical comparability. *International Journal of Research & Method in Education*, *36*(4), 415–438. https://doi.org/10.1080/1743727X.2012.710503

Asparouhov, T., & Muthén, B. O. (2010). *Weighted least squares estimation with missing data. Technical appendix.*

Assor, A., Kaplan, H., Kanat-Maymon, Y., & Roth, G. (2005). Directly controlling teacher behaviors as predictors of poor motivation and engagement in girls and boys: The role of anger and anxiety. *Learning and Instruction*, *15*(5), 397–413. https://doi.org/10.1016/J.LEARNINSTRUC.2005.07.008

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *The British Journal of Educational Psychology*, *72*(Pt 2), 261–278. https://doi.org/10.1348/000709902158883

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*(6, Pt.1), 359–372. https://doi.org/10.1037/h0043445

Atkinson, J. W. (1958). *Motives in fantasy, action, and society: a method of assessment and study*. Van Nostrand.

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. Wiley.

Augustine, A. A., & Larsen, R. J. (2012). Is a Trait Really the Mean of States? Similarities and Differences Between Traditional and Aggregate Assessments of Personality [Article]. *Journal of Individual Differences*, *33*(3), 131–137. https://doi.org/10.1027/1614-0001/a000083

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall.

Bandura, A. (2000). Exercise of Human Agency through Collective Efficacy [Article]. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age.

Barron, K. E., & Hulleman, C. S. (2015). Expectancy-Value-Cost Model of Motivation. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 503–509). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26099-6

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., Macleod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating Personality Structure, Personality Process, and Personality Development [Article]. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Beal, D. J. (2015). ESM 2.0: State of the Art and Future Potential of Experience Sampling Methods in Organizational Research [Article]. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335

Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, *71*(2), 287–311. https://doi.org/10.1093/poq/nfm006.

Benson, J. (1998). Developing a Strong Program of Construct Validation: A Test Anxiety Example. *Educational Measurement: Issues and Practice*, *17*(1), 10–17. https://doi.org/10.1111/j.1745-3992.1998.tb00616.x

Berry, J. W. (1969). On Cross-Cultural Comparability. *International Journal of Psychology*, *4*(2), 119–128. https://doi.org/10.1080/00207596908247261

Berry, J. W. (1989). Imposed Etics - Emics - Derived Etics: The Operationalization of A Compelling Idea. *International Journal of Psychology*, *24*(6), 721–735. https://doi.org/10.1080/00207598908247841

Berry, J. W. (1990). Imposed etics, emics, derived etics: Their conceptual and operational status in cross-cultural psychology. In T. N. Headland, K. L. Pike, & M. Harris (Eds.), *Emics and etics: The insider/outsider debate* (pp. 28–47). Sage.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: research and applications*. Cambridge University Press.

Bieg, M., Götz, T., & Lipnevich, A. A. (2014). What Students Think They Feel Differs from What They Really Feel – Academic Self-Concept Moderates the Discrepancy between Students' Trait and State Emotional Self-Reports. *PLoS ONE*, *9*(3), e92563. https://doi.org/10.1371/journal.pone.0092563.

Blair, E., & Burton, S. (1987). Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions [Article]. *The Journal of Consumer Research*, *14*(2), 280–288. https://doi.org/10.1086/209112

Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in Math and Science: Do Mothers' Beliefs Matter 12 Years Later? *Journal of Educational Psychology*, *96*(1), 97–109. https://doi.org/10.1037/0022-0663.96.1.97

Boekaerts, M. (1996). Personality and the psychology of learning. *European Journal of Personality*, *10*(5), 377–404. https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<377::AID-PER261>3.0.CO;2-N

Bofah, E. A., & Hannula, M. S. (2015). TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties [Article]. *Large-Scale Assessments in Education*, *3*(1), 1–36. https://doi.org/10.1186/s40536-015-0014-y

Bogdan, R., & Biklen, S. K. (2007). *Qualitative research for education: an introduction to theory and methods* (5th ed.). Pearson A & B.

Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, *65*, 101250. https://doi.org/10.1016/j.learninstruc.2019.101250

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: an introduction to diary and experience sampling research*. Guilford Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons, Inc.

Bong, M. (1998). Tests of the Internal/External Frames of Reference Model With Subject-Specific Academic Self-Efficacy and Frame-Specific Academic Self-Concepts [Article]. *Journal of Educational Psychology*, *90*(1), 102–110. https://doi.org/10.1037/0022-0663.90.1.102

Bong, M., & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? [Article]. *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Borman, G. D., & u.a. (2003). Comprehensive school reform and achievement: A meta-analysis [Article]. *Review of Educational Research*, *73*(2), 125–230. https://doi.org/10.3102/00346543073002125

Bradby, H. (2002). Translating culture and language: a research note on multilingual settings. *Sociology of Health & Illness*, *24*(6), 842–855. https://doi.org/10.1111/1467-9566.00321.

Breil, S. M., Schweppe, P. C., Geukes, K., Biesanz, J. C., Quintus, M., Wagner, J., Wrzus, C., Nestler, S., & Back, M. D. (2022). The Incremental Validity of Average States: A Replication and Extension of Finnigan and Vazire (2018) [Article]. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000408

Bremner, N. (2021). The multiple meanings of “student-centred” or “learner-centred” education, and the case for a more flexible approach to defining it [Article]. *Comparative Education*, *57*(2), 159–186. https://doi.org/10.1080/03050068.2020.1805863

Bremner, N., Sakata, N., & Cameron, L. (2022). The outcomes of learner-centred pedagogy: A systematic review. *International Journal of Educational Development*, *94*.

British Psychological Society. (1992). *Ethical Principles for Research with Human Participants*.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. The Guilford Press.

Brownlee, J., Berthelsen, D., & Boulton-Lewis, G. (2004). Working with toddlers in child care: Personal epistemologies and practice. *European Early Childhood Education Research Journal*, *12*(1), 55–70. https://doi.org/10.1080/13502930485209311

Bryant, F. B. (2002). Assessing the Validity of Measurement. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding More Multivariate Statistics* (pp. 99–146). American Psychological Association.

Bulla, Dominik, Uwitonze, A. G., Malmberg, L.-E., & Strand, S. (2022). *The cognitive-affective structure of “icyizere” (confidence) and “ishyaka” (motivation) among Rwandan students. A mixed-methods study (submitted).*

Bulla, D., Malmberg, L.-E., & Strand, S. (2022a). *Does time-allocation across classroom activities predict motivational states? Findings from Rwandan secondary schools (submitted).*

Bulla, D., Malmberg, L.-E., & Strand, S. (2022b). *Testing the Generalized Internal/External Frame of Reference Model within Rwanda's Basic-Education and Boarding Schools. A Multigroup SEM Analysis (unpublished material).*

Bulla, D., Malmberg, L.-E., & Strand, S. (2022c). *The generalized internal/external frame of reference model with averages of self-reported motivational states. Findings from Rwanda. (submitted).*

Burdett, N., & James, Z. (2018). *Learning Achievement in Rwandan Schools (LARS) III. Primary 6 (P6) and Secondary 3 (S3) results report. High Quality Technical Assistance for Results (HEART). Oxford, UK.*

Byrne, B. M. (1984). The General/Academic Self-Concept Nomological Network: A Review of Construct Validation Research [Article]. *Review of Educational Research*, *54*(3), 427–456. https://doi.org/10.3102/00346543054003427

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171– 246). Rand McNally.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Carrol, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723–733.

Carter, E., Rose, P., Ani-Asamoah, B., Leonard, P., & Colenbrander, K. (2020). *Video recordings of classroom observations: Using the Teach tool in Rwandan secondary schools. Research and Policy Paper No. 20/9.* https://www.educ.cam.ac.uk/centres/real/publications/LIT_classroom obs_Rwanda_final_upload.pdf

Castro-Alvarez, S., Tendeiro, J. N., Meijer, R. R., & Bringmann, L. F. (2021). Using structural equation modeling to study traits and states in intensive longitudinal data [Article]. *Psychological Methods*, *27*(1), 17–43. https://doi.org/10.1037/met0000393

Chao, R. C.-L., & Lambert, K. (2013). Emic-Etic Approach. In *The Encyclopedia of Cross-Cultural Psychology* (pp. 469–471). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118339893.wbeccp190

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries [Article]. *Journal of Economic Perspectives*, *20*(1), 91–116. https://doi.org/10.1257/089533006776526058

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, F. M., van de Vijver, F. J. R., & Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. *American Psychologist*, *66*(7), 593–603. https://doi.org/10.1037/a0022389

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance [Article]. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, C., & Hong, Y. (2007). Cultural processes: Basic principles. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 785–804). The Guilford Press.

Chiu, M.-S. (2008). Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries [Article]. *Educational Research and Evaluation*, *14*(3), 235–254. https://doi.org/10.1080/13803610802048858

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural Variation in Eye Movements during Scene Perception [Article]. *Proceedings of the National Academy of Sciences - PNAS*, *102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Clark, D. A., & Bowles, R. P. (2018). Model Fit and Item Factor Analysis: Overfactoring, Underfactoring, and a Program to Guide Interpretation. *Multivariate Behavioral Research*, *53*(4), 544–558. https://doi.org/10.1080/00273171.2018.1461058

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Clayton, K. E., & Zusho, A. (2016). A cultural heuristic approach to the study of Jamaican undergraduate students' achievement motivation. *British Journal of Educational Psychology*, *86*(1), 8–36. https://doi.org/10.1111/bjep.12081.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Conn, K. M. (2017). Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Impact Evaluations. *Review of Educational Research*, *87*(5), 863–898. https://doi.org/10.3102/0034654317712025

Connor, C. M., Lara J, J., Crowe, E. C., & Meadows, J. G. (2009). Instruction, Student Engagement, and Reading Skill Growth in Reading First Classrooms [Article]. *The Elementary School Journal*, *109*(3), 221–250. https://doi.org/10.1086/592305

Cornelius-White, J. (2007). Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis [Article]. *Review of Educational Research*, *77*(1), 113–143. https://doi.org/10.3102/003465430298563

Creswell, J. W. (2007). *Qualitative inquiry & research design: choosing among five approaches* (Fourth). SAGE.

Crocker, L. M., & Algina, J. (2006). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests [Article]. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

D'Haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: illustrating its surplus value in educational effectiveness research [Article]. *School Effectiveness and School Improvement*, *21*(2), 209–235. https://doi.org/10.1080/09243450903581218

Deci, E. L. (1992). On the nature and function of motivational theories. *Psychological Science*, *3*(3), 167–171.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior [Article]. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DeJaeghere, J. (2018). Girls' educational aspirations and agency: imagining alternative futures through schooling in a low-resourced Tanzanian community. *Critical Studies in Education*, *59*(2), 237–255. https://doi.org/10.1080/17508487.2016.1188835.

DeStefano, J., Ralaingita, W., Costello, M., Sax, A., & Frank, A. (2012). *Task Order 7 Early Grade Reading and Mathematics in Rwanda. Final Report*.

DeVellis, R. F. (2017). *Scale development: theory and applications*. SAGE.

Dietrich, J., Moeller, J., Guo, J., Viljaranta, J., & Kracke, B. (2019). In-the-moment profiles of expectancies, task values, and costs [Article]. *Frontiers in Psychology*, *10*(JULY), 1662–1662. https://doi.org/10.3389/fpsyg.2019.01662

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction*, *47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009

Dimitrov, D. M. (2010). Testing for Factorial Invariance in the Context of Construct Validation [Article]. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Dobie, T. E. (2019). Expanding conceptions of utility: middle school students' perspectives on the usefulness of mathematics. *Mathematical Thinking and Learning*, *21*(1), 28–53. https://doi.org/10.1080/10986065.2019.1564969

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*(2), 309–326. https://doi.org/10.1111/j.2044-8317.1994.tb01039.x

Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology*, *70*(4), 519–538. https://doi.org/10.1348/000709900158281

Dowson, M., & McInerney, D. M. (2003). What do students say about their motivational goals? Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology*, *28*(1), 91–113. https://doi.org/10.1016/S0361-476X(02)00010-3

Drisko, J. W., & Maschi, T. (2016). *Content analysis*. Oxford University Press.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: how individual interest moderates the effects of situational factors on task interest. [Article]. *Journal of Educational Psychology*, *99*(3), 597. https://doi.org/10.1037/0022-0663.99.3.597

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A.J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 105–121). Guildford Press.

Eccles, J. S. (2009). Who Am I and What Am I Going to Do With My Life? Personal and Collective Identities as Motivators of Action. *Educational Psychologist*, *44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). W. H. Freeman.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *The American Psychologist*, *48*(2), 90–101. http://www.ncbi.nlm.nih.gov/pubmed/8442578

Eccles, J. S., Tonks, S. M., & Wigfield, A. (2004). Expectancy Value Theory in Cross-Cultural Perspective. In Gregory Arief D. Liem & Dennis M. McInerney (Eds.), *Big Theories Revisited*. Information Age Publishing.

Eccles, J. S., & Wigfield, A. (1995). In the Mind of the Actor: The Structure of Adolescents' Achievement Task Values and Expectancy-Related Beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215–225. http://journals.sagepub.com/doi/pdf/10.1177/0146167295213003

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 1–13.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. C. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development*, *64*(3), 830. https://doi.org/10.2307/1131221

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional and personality development* (pp. 1017–1094). Wiley.

Elliot, Andrew J., & Hulleman, C. S. (2017). Achievement Goals. In Andrew J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of Competence and Motivation, Second Edition : Theory and Application* (pp. 43–60).

Elliot, Andrew J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010a). Cross-National Patterns of Gender Differences in Mathematics [Article]. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010b). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Erickson, J. D. (2019). (VAL) Primary Students' Emic Views of Reading Intervention: A Qualitative Case Study of Motivation. *Literacy Research: Theory, Method, and Practice*, *68*(1), 86–107. https://doi.org/10.1177/2381336919870282

Eynde, P. O. 't, De Corte, E., & Verschaffel, L. (2002). Framing Students' Mathematics-Related Beliefs. In *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 13–37). Springer Netherlands. https://doi.org/10.1007/0-306-47958-3_2

Fabb, N. (2005). *Sentence structure* (2nd ed.). Routledge.

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The Eye-Mind Wandering Link: Identifying Gaze Indices of Mind Wandering Across Tasks [Article]. *Journal of Experimental Psychology. Human Perception and Performance*, *46*(10), 1201–1221. https://doi.org/10.1037/xhp0000743

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Ferguson, T. D., Briesch, A. M., Volpe, R. J., Donaldson, A. R., & Feinberg, A. B. (2018). Psychometric Considerations for Conducting Observations Using Time-Sampling Procedures. *Assessment for Effective Intervention*, *44*(1), 45–57. https://doi.org/10.1177/1534508417747389

Finch, W. H. (2020). Using Fit Statistic Differences to Determine the Optimal Number of Factors to Retain in an Exploratory Factor Analysis [Article]. *Educational and Psychological Measurement*, *80*(2), 217–241. https://doi.org/10.1177/0013164419865769

Finegan, E. (2004). *Language: its structure and use* (4th ed.). Harcourt Brace College Publishers.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and Inattentive-Withdrawn Behavior and Achievement among Fourth Graders. *The Elementary School Journal*, *95*(5), 421–434. https://doi.org/10.1086/461853

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: a second course* (pp. 439–492). https://works.bepress.com/sara_finney/1/

Finnigan, K. M., & Vazire, S. (2018). The Incremental Validity of Average State Self-Reports Over Global Self-Reports of Personality [Article]. *Journal of Personality and Social Psychology*, *115*(2), 321–337. https://doi.org/10.1037/pspp0000136

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: the reasoned action approach*. Psychology Press.

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual Environment, Attention Allocation, and Learning in Young Children: When too much of a good thing may be bad. *Psychological Science*, *25*(7), 1362–1370. https://doi.org/10.1177/0956797614533801

Flake, J. K., Barron, K. E., Hulleman, C. S., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Fleeson, W. (2001). Toward a Structure- and Process-Integrated View of Personality: Traits as density distribution of states [Article]. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2009). On the Difference Between Experience-Sampling Self-Reports and Other Self-Reports [Article]. *European Journal of Personality*, *23*(5), 410–412.

Fleeson, W., & Gallagher, P. (2009). The Implications of Big Five Standing for the Distribution of Trait Manifestation in Behavior [Article]. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation [Article]. *Canadian Journal of Behavioural Science*, *49*(2), 78–88. https://doi.org/10.1037/cbs0000069

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299.

https://doi.org/10.1037/1040-3590.7.3.286

Fontaine, J. R. J., Poortinga, Y. H., Setiadi, B., & Markam, S. S. (2002). Cognitive structure of emotion terms in Indonesia and The Netherlands. *Cognition & Emotion*, *16*(1), 61–86. https://doi.org/10.1080/02699933014000130.

Fontaine, J. R. J., Scherer, K. R., & Soriano, C. (2013). *Components of emotional meaning: a sourcebook*.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence [Article]. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fredricks, J. A., Wang, M.-T., Schall Linn, J., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement [Article]. *Learning and Instruction*, *43*, 5–15. https://doi.org/10.1016/j.learninstruc.2016.01.009

Frenzel, A. C., Götz, T., Pekrun, R. H., & Watt, H. M. G. (2010). Development of Mathematics Interest in Adolescence: Influences of Gender, Family, and School Context. *Journal of Research on Adolescence*, *20*(2), 507–537. https://doi.org/10.1111/j.1532-7795.2010.00645.x.

Fridhandler, B. M. (1986). Conceptual Note on State, Trait, and the State-Trait Distinction [Article]. *Journal of Personality and Social Psychology*, *50*(1), 169–174. https://doi.org/10.1037/0022-3514.50.1.169

Frost, M., & Little, A. W. (2014). Children's learning practices in Ethiopia: observations from primary school classes [Article]. *Oxford Review of Education*, *40*(1), 91–111. https://doi.org/10.1080/03054985.2013.873526

Fuller, B., & Snyder, C. W. (1991). Vocal Teachers, Silent Pupils? Live in Botswana Classrooms. *Comparative Education Review*, *35*(2).

Fuller, K. A., Karunaratne, N. S., Naidu, S., Exintaris, B., Short, J. L., Wolcott, M. D., Singleton, S., & White, P. J. (2018). Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PLOS ONE*, *13*(10), e0205828. https://doi.org/10.1371/journal.pone.0205828

Garn, A. C., Simonton, K. L., Dasingert, T., & Simonton, A. (2017). Predicting changes in student engagement in university physical education: Application of control-value theory of achievement emotions. *Psychology of Sport and Exercise*, *29*, 93–102. https://doi.org/10.1016/j.psychsport.2016.12.005

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*(4), 454–474. https://doi.org/10.1037/a0030005

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via monte carlo simulation. *Psychological Methods*, *21*(1), 93–111. https://doi.org/10.1037/met0000064

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. In *International Journal of Nursing Studies* (Vol. 51, Issue 3, pp. 511–521). https://doi.org/10.1016/j.ijnurstu.2013.10.005

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, *107*(3), 663–677. https://doi.org/10.1037/edu0000003

Gaspard, H., Lauermann, F., Rose, N., Wigfield, A., & Eccles, J. S. (2020). Cross-Domain Trajectories of Students' Ability Self-Concepts and Intrinsic Values in Math and Language Arts [Article]. *Child Development*, *91*(5), 1800–1818. https://doi.org/10.1111/cdev.13343

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, *52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.003

Geertz, C. (1973). *The interpretation of cultures: selected essays*. Basic Books.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure Twice, Cut Down Error: A Process for Enhancing the Validity of Survey Scales [Article]. *Review of General Psychology*, *15*(4), 380–387. https://doi.org/10.1037/a0025704

Geiser, C. (2013). *Data analysis with Mplus*. Guilford Press.

Geiser, C. (2020). *Longitudinal Structural Equation Modeling with Mplus: A Latent State-Trait Perspective*. Guilford Press.

Geiser, C., Bishop, J., Lockhart, G., Shiffman, S., & Grenard, J. L. (2013). Analyzing Latent State-Trait and Multiple-Indicator Latent Growth Curve Models as Multilevel Structural Equation Models [Article]. *Frontiers in Psychology*, *4*, 975. https://doi.org/10.3389/fpsyg.2013.00975

Gelman, A. (2014). *Bayesian data analysis* (Third edit).

Gobel, P. (2008). Student Off-task Behavior and Motivation in the CALL Classroom [Article]. *International Journal of Pedagogies & Learning*, *4*(4), 4–18. https://doi.org/10.5172/ijpl.4.4.4

Godwin, K. E., Almeda, M. V., Petroccia, M., Baker, R. S. J. D., & Fisher, A. V. (2013). Classroom activities and off-task behavior in elementary school children. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.* (pp. 2428–2433). https://escholarship.org/uc/item/8mx9h5hq

Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S. J. D., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, *44*, 128–143. https://doi.org/10.1016/j.learninstruc.2016.04.003

Godwin, K. E., & Fisher, A. V. (2011). *Allocation of Attention in Classroom Environments: Consequences for Learning*. https://mindmodeling.org/cogsci2011/papers/0643/paper0643.pdf

Godwin, K. E., Seltman, H., Almeda, M., Davis Skerbetz, M., Kai, S., Baker, R. S., & Fisher, A. V. (2021). The elusive relationship between time on-task and learning: not simply an issue of measurement [Article]. *Educational Psychology (Dorchester-on-Thames)*, *41*(4), 502–519. https://doi.org/10.1080/01443410.2021.1894324

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley.

Gorsuch, R. (2015). *Factor analysis*. Routledge. http://files/755/search.html

Götz, T., Bieg, M., Luedtke, O., Pekrun, R. H., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? [Article]. *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Götz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. H. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33*(1), 9–33. https://doi.org/10.1016/j.cedpsych.2006.12.002.

Götz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion*, *34*(1), 49–62. https://doi.org/10.1007/s11031-009-9152-2.

Götz, T., Hall, N. C., Frenzel, A. C., & Pekrun, R. H. (2006). A hierarchical conceptualization of enjoyment in students. *Learning and Instruction*, *16*(4), 323–338. https://doi.org/10.1016/J.LEARNINSTRUC.2006.07.004.

Götz, T., Sticca, F., Pekrun, R. H., Murayama, K., & Elliot, A. J. (2016). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction*, *41*, 115–125. https://doi.org/10.1016/J.LEARNINSTRUC.2015.10.007.

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of Not Interpreting Structure Coefficients in Published CFA Research: A Reminder [Article]. *Structural Equation Modeling*, *10*(1), 142–153. https://doi.org/10.1207/S15328007SEM1001_7

Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World [Article]. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Grant, H., & Dweck, C. S. (2003). Clarifying Achievement Goals and Their Impact. *Journal of Personality and Social Psychology*, *85*(3), 541–553. https://doi.org/10.1037/0022-3514.85.3.541

Griskell, H. L., Gámez, P. B., & Lesaux, N. K. (2020). Assessing middle school dual language learners' and English-only students' motivation to participate in classroom discussion. *Learning and Individual Differences*, *77*, 101799. https://doi.org/10.1016/j.lindif.2019.101799

Guillaume, D. W., & Khachikian, C. S. (2011). The effect of time-on-task on student grades and grade expectations. *Assessment & Evaluation in Higher Education*, *36*(3), 251–261. https://doi.org/10.1080/02602930903311708

Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data [Article]. *Learning and Instruction*, *58*, 210–219. https://doi.org/10.1016/j.learninstruc.2018.07.007

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction*, *49*, 81–91. https://doi.org/10.1016/J.LEARNINSTRUC.2016.12.007

Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional Contexts for Engagement and Achievement in Reading. In *Handbook of Research on Student Engagement* (pp. 601–634). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_29

Haddad, W. D., Colletta, N. J., Fisher, N., Lakin, M., & Sutton, M. (1990). *Meeting basic learning needs: A vision for the 1990s. Background Document World Conference on Education for All Meeting Basic Learning Needs Jomtien, Thailand 5-9 March 1990.*

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: a structural model [Article]. *Educational Psychology (Dorchester-on-Thames)*, *28*(1), 59–71. https://doi.org/10.1080/01443410701413753

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The

Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, *111*(5), 745–765.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012a). Helping Parents to Motivate Adolescents in Mathematics and Science: An Experimental Test of a Utility-Value Intervention [Article]. *Psychological Science*, *23*(8), 899–906. https://doi.org/10.1177/0956797611435530

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012b). Helping Parents to Motivate Adolescents in Mathematics and Science. *Psychological Science*, *23*(8), 899–906. https://doi.org/10.1177/0956797611435530

Harter, S. (2006). The Self. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology (Vol. 3)* (pp. 505–570). Wiley.

Hattie, J. A. (2009). *Visible learning : a synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods Introduction to Content Validity. *Psychological Assessment*, *7*(3), 238–247. http://personal.kent.edu/~dfresco/CRM_Readings/Haynes_1995.pdf

Heck, R. H., & Thomas, S. L. (2020). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus*. https://doi.org/10.4324/9781410604767

Heine, S. J. (2001). Self as Cultural Product: An Examination of East Asian and North American Selves [Article]. *Journal of Personality*, *69*(6), 881–905. https://doi.org/10.1111/1467-6494.696168

Heine, S. J. (2003). An exploration of cultural variation in self-enhancing and self-improving motivations. In V. Murphy-Berman & J. J. Berman (Eds.), *Nebraska symposium on motivation: Cross-cultural differences in perspectives on the self. Vol. 49* (pp. 101–128). University of Nebraska Press.

Heine, S. J. (2005). Constructing good selves in Japan and North America. In R. M. Sorrentino, D. Cohen, J. M. Olson, & M. P. Zanna (Eds.), *Culture and Social Behavior: The Tenth Ontario Symposium* (pp. 95–116). Lawrence Erlbaum.

Heine, S. J. (2016). *Cultural psychology*.

Helm, F., Mueller-Kalthoff, H., Nagy, N., & Möller, J. (2016). Dimensional Comparison Theory: Perceived Subject Similarity Impacts on Students' Self-Concepts [Article]. *AERA Open*, *2*(2), 233285841665062. https://doi.org/10.1177/2332858416650624

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision, and action: eye movements and the visual world*. Psychology Press. http://solo.bodleian.ox.ac.uk/primo_library/libweb/action/display.do?tabs=detailsTab&ct=display&fn=search&doc=oxfaleph000913430&indx=1&recIds=oxfaleph000913430&recIdxs=0&elementId=0&renderMode=poppedOut&displayMode=full&frbrVersion=&frbrSourceidDisplay=ox

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? [Article]. *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice [Article]. *Educational and*

Psychological Measurement, *66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Hidi, S. (1990). Interest and Its Contribution as a Mental Resource for Learning [Article]. *Review of Educational Research*, *60*(4), 549–571. https://doi.org/10.3102/00346543060004549

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Higgins, J. P. T. (2019). *Cochrane handbook for systematic reviews of interventions* (Second edi).

Hill, M. (1997). Participatory research with children. *Child & Family Social Work*, *2*(3), 171–183. https://doi.org/10.1046/j.1365-2206.1997.00056.x

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967–988. https://doi.org/10.1016/0149-2063(95)90050-0

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, *1*(1), 104–121. https://doi.org/10.1177/109442819800100106.

Hintze, J. M. (2005). Psychometrics of Direct Observation [Article]. *School Psychology Review*, *34*(4), 507–519. https://doi.org/10.1080/02796015.2005.12088012

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: software of the mind*. McGraw-Hill.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, *44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A Comparison of ML, WLSMV, and Bayesian Methods for Multilevel Structural Equation Models in Small Samples: A Simulation Study [Article]. *Multivariate Behavioral Research*, *51*(5), 661–680. https://doi.org/10.1080/00273171.2016.1208074

Honeyman, C. A. (2015a). *Early Literacy Promotion in Rwanda: Opportunities and Obstacles Based on a Study Conducted for Save The Children*. https://rwanda.savethechildren.net/sites/rwanda.savethechildren.net/files/library/Early Literacy Promotion Rwanda_OpportunitiesObstacles_2015-01-27.pdf

Honeyman, C. A. (2015b). *The orderly entrepreneur: youth, education, and governance in Rwanda*.

Hospel, V., Galand, B., & Janosz, M. (2016). Multidimensionality of behavioural engagement: Empirical support and implications [Article]. *International Journal of Educational Research*, *77*, 37–49. https://doi.org/10.1016/j.ijer.2016.02.007

Howe, C., & Abedin, M. (2013). Classroom dialogue: a systematic review across four decades of research [Article]. *Cambridge Journal of Education*, *43*(3), 325–356. https://doi.org/10.1080/0305764X.2013.786024

Hox, J. J., & Maas, C. J. M. (2001). The Accuracy of Multilevel Structural Equation Modeling With Pseudobalanced Groups and Small Samples [Article]. *Structural Equation Modeling*, *8*(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1

Hox, J. J., Van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/J.JSP.2011.07.001

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*(3), 422–449. https://doi.org/10.1037/a0018947

Hutt, S. J., & Hutt, C. (1970). *Direct observation and measurement of behavior*. Charles C. Thomas.

Ireson, J., & Hallam, S. (2001). *Ability grouping in education*. Sage.

Irwing, P., Booth, T., & Hughes, D. J. (2018). *The Wiley handbook of psychometric testing: a multidisciplinary reference on survey, scale and test development* (1st ed.).

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, *63*(6), 932–944. http://www.ncbi.nlm.nih.gov/pubmed/1460561

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, *102*(3), 588–600. https://doi.org/10.1037/a0019682

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. https://doi.org/10.1016/J.CEDPSYCH.2014.11.002

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and Assimilation Effects of Dimensional Comparisons in Five Subjects: An Extension of the I/E Model [Article]. *Journal of Educational Psychology*, *107*(4), 1086–1101. https://doi.org/10.1037/edu0000021

Ji, Y., Luo, W., Lai, M. H. ., Yoon, M., Chen, L.-S., & Kwok, O.-M. (2021). Model-Selection-Based Approaches to Identifying the Optimal Number of Factors in Multilevel Exploratory Factor Analysis [Article]. *Structural Equation Modeling*, *28*(5), 763–777. https://doi.org/10.1080/10705511.2021.1916394

Johnson, T., & van de Vijver, F. J. R. (2003). Social desirability in crosscultural research. In J. Harkness, F. J. R. van de Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195–206). Wiley.

Jones, S., Schipper, Y., Ruto, S., & Rajani, R. (2014). Can Your Child Read and Count? Measuring Learning Outcomes in East Africa [Article]. *Journal of African Economies*, *23*(5), 643–672. https://doi.org/10.1093/jae/eju009

Jonker, C., Mojaki, L., Meiring, D., & Fontaine, J. R. J. (2013). Adaptation of the GRID instrument in Setswana. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of Emotional Meaning: a Sourcebook.* (pp. 455–461).

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., Groot, E.

de, Gilbert, M. C., Musu, L., Kempler, T. M., & Kelly, K. L. (2007). Cognitive Processing of Self-Report Items in Educational Research: Do They Think What We Mean? *Educational Psychologist*, *42*(3), 139–151. https://doi.org/10.1080/00461520701416231

Karasawa, M., Little, T. D., Miyashita, T., Mashima, M., & Azuma, H. (1997). Japanese Children's Action-Control Beliefs about School Performance [Article]. *International Journal of Behavioral Development*, *20*(3), 405–423. https://doi.org/10.1080/016502597385207

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel Factor Analysis: Reporting Guidelines and a Review of Reporting Practices [Article]. *Multivariate Behavioral Research*, *51*(6), 881–898. https://doi.org/10.1080/00273171.2016.1228042

Kimenyi, A. (1978). *A relational grammar of Kinyarwanda*. University of California Press.

King, E. (2014). *From classrooms to conflict in Rwanda*.

King, R. B., & McInerney, D. M. (2014). Culture's Consequences on Student Motivation: Capturing Cross-Cultural Universality and Variability Through Personal Investment Theory. *Educational Psychologist*, *49*(3), 175–198. https://doi.org/10.1080/00461520.2014.926813.

King, R. B., & McInerney, D. M. (2016). Culturalizing motivation research in educational psychology. *British Journal of Educational Psychology*, *86*(1), 1–7. https://doi.org/10.1111/bjep.12106

King, R. B., & McInerney, D. M. (2019). Family-support goals drive engagement and achievement in a collectivist context: Integrating etic and emic approaches in goal research. *Contemporary Educational Psychology*, *58*, 338–353. https://doi.org/10.1016/J.CEDPSYCH.2019.04.003

King, R. B., McInerney, D. M., & Pitliya, R. J. (2018). Envisioning a Culturally Imaginative Educational Psychology. *Educational Psychology Review*, *30*(3), 1031–1065. https://doi.org/10.1007/s10648-018-9440-z.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency [Article]. *Computers in Human Behavior*, *25*(2), 306–314. https://doi.org/10.1016/j.chb.2008.12.008

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching [Article]. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kitayama, S., & Cohen, D. (2007). *Handbook of cultural psychology*. Guilford Press.

Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective [Article]. *International Journal of Psychology*, *39*(3), 205–230. https://doi.org/10.1080/00207590344000330

Kline, R. B. (2013). Reverse Arrow Dynamics: Feedback Loops and Formative Measurement. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 39–77). http://psychology.concordia.ca/fac/kline/library/k12.pdf

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, *11*(4), 415–425. https://doi.org/10.1037/1040-3590.11.4.415.

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A Practical Measure of Student Motivation: Establishing Validity Evidence for the Expectancy-Value-Cost Scale in Middle School. *The Journal of Early Adolescence*, *35*(5–6), 790–816. https://doi.org/10.1177/0272431614556890

Kreft, I., & Leeuw, J. de. (1998). *Introducing Multilevel Modeling*. Sage Publications, Inc. http://gifi.stat.ucla.edu/janspubs/1998/books/kreft_deleeuw_B_98.pdf

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis [Article]. *Educational Research Review*, *25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Sage.

Krippendorff, K., & Craggs, R. (2016). The Reliability of Multi-Valued Coding of Data. *Communication Methods and Measures*, *10*(4), 181–198. https://doi.org/10.1080/19312458.2016.1228863

Krishnaratne, S., White, H., & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries. Paper 20.* (No. 20). www.3ieimpact.org

Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The Role of Time on Task in Computer-Based Low-Stakes Assessment of Cross-Curricular Skills [Article]. *Journal of Educational Psychology*, *106*(3), 627–638. https://doi.org/10.1037/a0035507

Lam, S. fong, Jimerson, S., Shin, H., Cefai, C., Veiga, F. H., Hatzichristou, C., Polychroni, F., Kikas, E., Wong, B. P. H., Stanculescu, E., Basnett, J., Duck, R., Farrell, P., Liu, Y., Negovan, V., Nelson, B., Yang, H., & Zollneritsch, J. (2016). Cultural universality and specificity of student engagement in school: The results of an international study from 12 countries. *British Journal of Educational Psychology*, *86*(1), 137–153. https://doi.org/10.1111/bjep.12079

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. http://www.ncbi.nlm.nih.gov/pubmed/843571

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy–value theory of achievement-related behaviors. *Developmental Psychology*, *53*(8), 1540–1559. https://doi.org/10.1037/dev0000367

Lavrijsen, J., Vansteenkiste, M., Boncquet, M., & Verschueren, K. (2021). Does motivation predict changes in academic achievement beyond intelligence and personality? A multitheoretical perspective. *Journal of Educational Psychology*. https://doi.org/10.1037/EDU0000666

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation Interventions in Education: A Meta-Analytic Review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries [Article]. *Learning and Individual Differences*, *19*(3), 355–365. https://doi.org/10.1016/j.lindif.2008.10.009

Lee, M., & Bong, M. (2016). In their own words: Reasons underlying the achievement striving of students in schools. *Journal of Educational Psychology*, *108*(2), 274–294. https://doi.org/10.1037/edu0000048

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis [Article]. *Social Behavior and Personality*, *46*(3), 517–528. https://doi.org/10.2224/sbp.7054

Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology*, *47*(1), 233–247. https://doi.org/10.1037/a0021307

Li, Y., & Lerner, R. M. (2013). Interrelations of Behavioral, Emotional, and Cognitive School Engagement in High School Students. *Journal of Youth and Adolescence*, *42*(1), 20–32.

https://doi.org/10.1007/s10964-012-9857-5

Linnenbrink-Garcia, L., Patall, E. A., & Messersmith, E. E. (2013). Antecedents and consequences of situational interest. *British Journal of Educational Psychology*, *83*(4), 591–614. https://doi.org/10.1111/j.2044-8279.2012.02080.x

Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2004). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context [Article]. *British Journal of Educational Psychology*, *75*(4), 567–586. https://doi.org/10.1348/000709905X42239

Loevinger, J. (1957). Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

Louw, J., Muller, J., & Tredoux, C. (2008). Time-on-task, technology and mathematics achievement [Article]. *Evaluation and Program Planning*, *31*(1), 41–50. https://doi.org/10.1016/j.evalprogplan.2007.11.001

Luo, W., Ng, P. T., Lee, K., & Aye, K. M. (2016). Self-efficacy, value, and achievement emotions as mediators between parenting practice and homework behavior: A control-value theory perspective. *Learning and Individual Differences*, *50*, 275–282. https://doi.org/10.1016/j.lindif.2016.07.017

Magrath, B., Aslam, M., & Johnson, D. (2019). Systems Research in Education: Designs and methods [Article]. *Research in Comparative and International Education*, *14*(1), 7–29. https://doi.org/10.1177/1745499919828927

Malmberg, L.-E., Walls, T. A., Martin, A. J., Little, T. D., & Lim, W. H. T. (2013). (4) Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and Individual Differences*, *28*, 54–65. https://doi.org/10.1016/J.LINDIF.2013.09.007

Marbell, K. N., & Grolnick, W. S. (2013). Correlates of parental control and autonomy support in an interdependent culture: A look at Ghana [Article]. *Motivation and Emotion*, *37*(1), 79–92. https://doi.org/10.1007/s11031-012-9289-2

Markus, H. R. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63–78. https://doi.org/10.1037/0022-3514.35.2.63

Markus, H. R., & Kitayama, S. (2010). Cultures and Selves: A Cycle of Mutual Constitution [Article]. *Perspectives on Psychological Science*, *5*(4), 420–430. https://doi.org/10.1177/1745691610375557

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model [Article]. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037//0022-0663.81.3.417

Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, *2*(2), 77–172. https://doi.org/10.1007/BF01322177

Marsh, H. W. (1999). *Self Description Questionnaire – II*.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. D. (2013). Factorial, Convergent, and Discriminant Validity of TIMSS Math and Science Motivation Measures: A Comparison of Arab and Anglo-Saxon Countries [Article]. *Journal of Educational Psychology*, *105*(1), 108–128. https://doi.org/10.1037/a0029907

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., Möller, J., & Abu-Hilal, M. M. (2015). The Internal/External Frame of Reference Model of Self-Concept and Achievement Relations: Age-Cohort and Cross-Cultural Differences. *American Educational Research Journal*, *52*(1), 168–202. https://doi.org/10.3102/0002831214549453

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology*, *38*(3), 376–393. http://www.ncbi.nlm.nih.gov/pubmed/12005381

Marsh, H. W., & Hau, K.-T. (1996). Assessing Goodness of Fit: Is Parsimony Always Desirable? [Article]. *The Journal of Experimental Education*, *64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604

Marsh, H. W., & Hau, K.-T. (2004). Explaining Paradoxical Relations Between Academic Self-Concepts and Achievements: Cross-Cultural Generalizability of the Internal/External Frame of Reference Predictions Across 26 Countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal Multilevel Models of the Big-Fish-Little-Pond Effect on Academic Self-Concept [Article]. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037/0022-3514.78.2.337

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and non-native languages for Chinese students. [Article]. *Journal of Educational Psychology*, *93*(3), 543–553. https://doi.org/10.1037/0022-0663.93.3.543

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation [Article]. *Contemporary Educational Psychology*, *39*(4), 326–341. https://doi.org/10.1016/j.cedpsych.2014.08.003

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional Comparison Theory: Paradoxical relations between self-beliefs and achievements in multiple domains [Article]. *Learning and Instruction*, *35*, 16–32. https://doi.org/10.1016/j.learninstruc.2014.08.005

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom Climate and Contextual Effects: Conceptual and Methodological Issues in the Evaluation of Group-Level Effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Pekrun, R. H., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., Seaton, M., Dicke, T., Parker, P. D., & Horwood, M. S. (2019). The Centrality of Academic Self-Concept to Motivation and Learning. In *The Cambridge Handbook of Motivation and Learning* (pp. 36–62).

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., Walker, R., & Debus, R. (1991). Subject-specific components of academic self-concept and self-efficacy. *Contemporary Educational Psychology*, *16*(4), 331–345. https://doi.org/10.1016/0361-476X(91)90013-B

Marsh, H. W., & Yeung, A. S. (1997). Causal Effects of Academic Self-Concept on Academic Achievement [Article]. *Journal of Educational Psychology*, *89*(1), 41–54. https://doi.org/10.1037/0022-0663.89.1.41

Marsh, H. W., & Yeung, A. S. (2001). An Extension of the Internal/External Frame of Reference Model: A Response to Bong [Article]. *Multivariate Behavioral Research*, *36*(3), 389–420. https://doi.org/10.1207/S15327906389-420

Marshall, L. R. (2016). 'Going to school to become good people': Examining aspirations to respectability and goodness among schoolchildren in urban Ethiopia. *Childhood*, *23*(3), 423–437. https://doi.org/10.1177/0907568216634065.

Martin, A. J., Papworth, B., Ginns, P., Malmberg, L.-E., Collie, R. J., & Calvo, R. A. (2015). Real-time motivation and engagement during a month at school: Every moment of every day for every student matters. *Learning and Individual Differences*, *38*, 26–35. https://doi.org/10.1016/j.lindif.2015.01.014

Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: conceptual and analytical issues. *Gerontology*, *50*(1), 7–11. https://doi.org/10.1159/000074382

Mboya, M. M. (1995). Perceived Teachers' Behaviours and Dimensions of Adolescent Self-concepts [Article]. *Educational Psychology (Dorchester-on-Thames)*, *15*(4), 491–499. https://doi.org/10.1080/0144341950150410

McAdams, D. P., & Pals, J. L. (2006). A New Big Five: Fundamental Principles for an Integrative Science of Personality [Article]. *The American Psychologist*, *61*(3), 204–217. https://doi.org/10.1037/0003-066X.61.3.204

McCoach, B. D., Gable, R. K., & Madura, J. P. (2014). *Instrument development in the affective domain: school and corporate applications.* Springer. https://www.amazon.co.uk/Instrument-Development-Affective-Domain-Applications-ebook/dp/B00CPHP62S/ref=sr_1_1?keywords=Instrument+development+in+the+affective+domain%3A+School+and+corporate+applications&qid=1568453127&sr=8-1

McCoy, D. C., Wolf, S., & Godfrey, E. B. (2014). Student motivation for learning in Ghana: Relationships with caregivers' values toward education, attendance, and academic achievement. *School Psychology International*, *35*(3), 294–308. https://doi.org/10.1177/0143034313508055.

McCrae, R. R., & Costa, P. T. J. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). Guilford Press.

McCrae, R. R., & Costa, P. T. J. (1999). A five-factor theory of personality. In L. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 139–153). Guilford Press.

McCrae, R. R., Costa, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., Sanz, J., Sánchez-Bernardos, M. L., Kusdil, M. E., Woodfield, R., Saunders, P. R., & Smith, P. B. (2000). Nature Over Nurture: Temperament, Personality, and Life Span Development [Article]. *Journal of Personality and Social Psychology*, *78*(1), 173–186. https://doi.org/10.1037/0022-3514.78.1.173

McKenzie, J., Wood, Kotecki, Clark, J., & Brey, R. (1999). Establishing Content Validity: Using Qualitative and Quantitative Steps. *American Journal of Health Behavior*, *23*, 311–318.

McNeish, D., & Hamaker, E. L. (2020). A Primer on Two-Level Dynamic Structural Equation Models for Intensive Longitudinal Data in Mplus [Article]. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the Unnecessary Ubiquity of Hierarchical Linear Modeling [Article]. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

McWilliam, E., Dooley, K., McArdle, F., & Tan, J. P.-L. (2009). Voicing objections. In A. Jackson & L. Mazzei (Eds.), *Voice in Qualitative Inquiry: Challenging Conventional, Interpretive, and Critical Conceptions in Qualitative Research.* (pp. 63–75). Routledge.

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. Guilford Press.

Merrell, K. W. (2013). Behavioral, social, and emotional assessment of children and adolescents, third edition. In *Behavioral Social and Emotional Assessment of Children and Adolescents, Third Edition* (5th editio). https://doi.org/10.4324/9780203827253

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: a guide to design and implementation* (Fourth). Jossey-Bass.

Mertens, D. M. (2010). *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). SAGE.

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning [Article]. *The American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Michener, C. J., Patrick Proctor, C., & Silverman, R. D. (2017). Features of instructional talk predictive of reading comprehension [Article]. *Reading & Writing*, *31*(3), 725–756. https://doi.org/10.1007/s11145-017-9807-4

Milligan, L. O., Clegg, J., & Tikly, L. (2016). Exploring the potential for language supportive learning in English medium instruction: a Rwandan case study. *Comparative Education*, *52*(3), 328–342. https://doi.org/10.1080/03050068.2016.1185258

MINECOFIN. (2000). *Rwanda Vision 2020*.

MINEDUC. (2015). *Competence Based Curriculum. Curriculum Framework Pre Primary to Upper Secondanry.*

MINEDUC. (2019). *2019 Education Statistics.*

Mischel, W., & Shoda, Y. (1995). A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure [Article]. *Psychological Review*, *102*(2), 246–268. https://doi.org/10.1037/0033-295X.102.2.246

Moffett, L., & Morrison, F. J. (2020). Off-Task Behavior in Kindergarten: Relations to Executive Function and Academic Achievement [Article]. *Journal of Educational Psychology*, *112*(5), 938–955. https://doi.org/10.1037/edu0000397

Möller, J., & Köller, O. (2001). Dimensional Comparisons: An Experimental Approach to the Internal/External Frame of Reference Model [Article]. *Journal of Educational Psychology*, *93*(4), 826–835. https://doi.org/10.1037/0022-0663.93.4.826

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The Generalized Internal/External Frame of Reference Model: An Extension to Dimensional Comparison Theory. *Frontline Learning Research*, *4*(2), 1–11. https://doi.org/10.14786/flr.v4i2.169

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains [Article]. *European Journal of Psychology of Education*, *21*(4), 467–487. https://doi.org/10.1007/BF03173515

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A Meta-Analysis of Relations Between Achievement and Self-Concept [Article]. *Review of Educational Research*, *90*(3), 376–419. https://doi.org/10.3102/0034654320919354

Montoya, A. K., & Edwards, M. C. (2021). The Poor Fit of Model Fit for Selecting Number of Factors in Exploratory Factor Analysis for Scale Evaluation [Article]. *Educational and Psychological Measurement*, *81*(3), 413–440. https://doi.org/10.1177/0013164420942899

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly Latent Multilevel Analyses of Classroom Climate: An Illustration [Article]. *The Journal of Experimental Education*, *82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the Effects of Classroom Discussion on Students' Comprehension of Text: A Meta-Analysis. *Journal of Educational Psychology*, *101*(3), 740–764. https://doi.org/10.1037/a0015576

Musker, P., Clist, P., Abbott, P., Boyd, C., & Latimer, K. (2014). *Evaluation of Results Based Aid in Rwandan Education – 2013 Evaluation Report.*

Muthén, B. O., & A., S. (1995). Complex ample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 267–316). American Sociological Association.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide. Eighth Edition.*

Nag, S. (2014). *Literacy, Foundation Learning and Assessment in Developing Countries*.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. In *Educational Research and Evaluation* (Vol. 12, Issue 4, pp. 323–345). https://doi.org/10.1080/13803610600765687

Neuman, W. L. (2014). *Social research methods : qualitative and quantitative approaches* (Seventh ed). Pearson Education Limited.

Nguyen, T. D., Cannata, M., & Miller, J. (2016). Understanding student behavioral engagement: Importance of student interaction with peers and teachers. *The Journal of Educational Research*. https://doi.org/10.1080/00220671.2016.1220359

Ngware, M. W., Mutisya, M., & Oketch, M. (2012). Patterns of teaching style and active teaching: Do they differ across subjects in low and high performing primary schools in Kenya? [Article]. *London Review of Education*, *10*(1), 35–54. https://doi.org/10.1080/14748460.2012.659058

Nicholls, J. G. (1979). Development of perception of own attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology*, *71*(1), 94–99. https://doi.org/10.1037/0022-0663.71.1.94

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*(4), 1170–1191. https://doi.org/10.1037/a0036307

Niepel, C., Stadler, M., & Greiff, S. (2019). Seeing Is Believing: Gender Diversity in STEM Is Related to Mathematics Self-Concept [Article]. *Journal of Educational Psychology*, *111*(6), 1119–1130. https://doi.org/10.1037/edu0000340

NISR. (2016). *Rwanda Demographic and Health Survey 2015*. https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf

Nzabalirwa, W. (2014). Rwanda: An Overview. In C. Wolhuter (Ed.), *Education in East and Central Africa* (pp. 299–320). Bloomsbury Academic.

O'Sullivan, M. (2004). The reconceptualisation of learner-centred approaches: a Namibian case study [Article]. *International Journal of Educational Development*, *24*(6), 585–602. https://doi.org/10.1016/S0738-0593(03)00018-X

Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press.

Ocumpaugh, J., Baker, R. S. J. D., Mercedes, M., Rodrigo, T., Baker, R., & Ocumpaugh, R. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual*. http://www.columbia.edu/~rsb2162/BROMP.pdf

OECD. (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. https://doi.org/10.1787/9789264190511-en

OECD. (2013b). *PISA 2012 results: Ready to learn—students' engagement, drive and self-beliefs (Vol. III)*.

Olivier, E., Galand, B., Hospel, V., & Dellisse, S. (2020). Understanding behavioural engagement and achievement: The roles of teaching practices and student sense of competence and task value. *British Journal of Educational Psychology*, *90*(4), 887–909. https://doi.org/10.1111/bjep.12342

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207

Orth, U., Dapp, L. C., Erol, R. Y., Krauss, S., & Luciano, E. C. (2020). Development of domain-specific self-evaluations: A meta-analysis of longitudinal studies [Article]. *Journal of Personality and Social Psychology*, *120*(1), 145–172. https://doi.org/10.1037/pspp0000378

Osborne, J. W. (2014). *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis* (Issue September).

Ozel, M., Caglak, S., & Erdogan, M. (2013). Are affective factors a good predictor of science achievement? Examining the role of affective factors based on PISA 2006 [Article]. *Learning and Individual Differences*, *24*, 73–82. https://doi.org/10.1016/j.lindif.2012.09.006

Parrisius, C., Gaspard, H., Zitzmann, S., Trautwein, U., & Nagengast, B. (2021). The "Situative Nature" of Competence and Value Beliefs and the Predictive Power of Autonomy Support: A Multilevel Investigation of Repeated Observations [Article]. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000680

Patall, E. A., Vasquez, A. C., Steingut, R. R., Trimble, S. S., & Pituch, K. A. (2016). Daily interest, engagement, and autonomy support in the high school science classroom. *Contemporary Educational Psychology*, *46*, 180–194. https://doi.org/10.1016/J.CEDPSYCH.2016.06.002

Patrick, A. W., & Zuckerman, M. (1977). An application of the state-trait concept to the need for achievement [Article]. *Journal of Research in Personality*, *11*(4), 459–465. https://doi.org/10.1016/0092-6566(77)90006-X

Paulsen, A., Overgaard, S., & Lauritsen, J. M. (2012). Quality of data entry using single entry, double entry and automated forms processing-an example based on a study of patient-reported outcomes. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0035087

Pearson, P. (2014). Policy without a plan: English as a medium of instruction in Rwanda [Article]. *Current Issues in Language Planning*, *15*(1), 39–56. https://doi.org/10.1080/14664208.2013.857286

Peixoto, F., Mata, L., Monteiro, V., Sanches, C., & Pekrun, R. H. (2015). The Achievement Emotions Questionnaire: Validation for Pre-Adolescent Students. *European Journal of Developmental Psychology*, *12*(4), 472–481. https://doi.org/10.1080/17405629.2015.1040757.

Pekrun, R. H. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9.

Pekrun, R. H., & Buehler, M. (2014). Self-Report Measures of Academic Emotions. In Reinhard Pekrun & Lisa Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 561–579). https://solo.bodleian.ox.ac.uk/shib/OX/interlib

Pekrun, R. H., Frenzel, A. C., Götz, T., & Perry, R. P. (2007). The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education. In *Emotion in Education* (pp. 13–36). Elsevier. https://doi.org/10.1016/B978-012372545-5/50003-4.

Pekrun, R. H., Götz, T., & Perry, R. P. (2005). *Achievement Emotions Questionnaire (AEQ): User's manual*.

Pekrun, R. H., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Götz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704.

Pekrun, R. H., & Linnenbrink-Garcia, L. (2014). *International handbook of emotions in education*. Routledge.

Pekrun, R. H., & Perry, R. P. (2014). Control-Value Theory of Achievement Emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 120–141). Routledge. https://doi.org/10.4324/9780203148211-12.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, *106*(1), 315–329. https://doi.org/10.1037/a0034027

Perez, T., Dai, T., Kaplan, A., Cromley, J. G., Brooks, W. D., White, A. C., Mara, K. R., & Balsai, M. J. (2019). Interrelations among expectancies, task values, and perceived costs in undergraduate biology achievement [Article]. *Learning and Individual Differences*, *72*, 26–38. https://doi.org/10.1016/j.lindif.2019.04.001

Peterson, P. L., Swing, S. R., Stark, K. D., & Waas, G. A. (1984). Students' Cognitions and Time on Task during Mathematics Instruction. *American Educational Research Journal*, *21*(3), 487. https://doi.org/10.2307/1162912

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*(1), 152–174. https://doi.org/10.1111/bjep.12028

Pitzer, J. R., & Skinner, E. A. (2017). Predictors of changes in students' motivational resilience over the school year. *International Journal of Behavioral Development*, *41*(1), 15–29. https://doi.org/10.1177/0165025416642051

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Lindell & Whit-Ney*. https://doi.org/10.1037/0021-9010.88.5.879.

Posti-Ahokas, H., & Palojoki, P. (2014). Navigating transitions to adulthood through secondary education: aspirations and the value of education for Tanzanian girls. *Journal of Youth Studies*, *17*(5), 664–681. https://doi.org/10.1080/13676261.2013.853871

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine [Article]. *Understanding Statistics*, *2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

Preckel, F., & Brüll, M. (2010). The benefit of being a big fish in a big pond: Contrast and assimilation effects on academic self-concept [Article]. *Learning and Individual Differences*, *20*(5), 522–531. https://doi.org/10.1016/j.lindif.2009.12.007

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly*, *68*(1), 109–130. https://doi.org/10.1093/poq/nfh008

Pritchett, L. (2013). *The rebirth of education: schooling ain't learning*. Center for Global Development.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect [Article]. *Intelligence (Norwood)*, *53*, 202–210. https://doi.org/10.1016/j.intell.2015.10.006

Punch, S. (2002). Research with Children. *Childhood*, *9*(3), 321–341. https://doi.org/10.1177/0907568202009003005

Putwain, D. W., Pekrun, R. H., Nicholson, L. J., Symes, W., Becker, S., & Marsh, H. W. (2018). Control-Value Appraisals, Enjoyment, and Boredom in Mathematics: A Longitudinal Latent Interaction

Analysis. *American Educational Research Journal*, 000283121878668. https://doi.org/10.3102/0002831218786689

Rabe-Hesketh, S., Skrondal, A., & Xiaahui, Z. (2012). Multilevel Structural Equation Models. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 512–531). Guildford Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Sage Publications.

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2020). Do Self-Reported Traits and Aggregated States Capture the Same Thing? A Nomological Perspective on Trait-State Homomorphy [Article]. *Social Psychological and Personality Science*, *10*(5), 596–611. https://doi.org/10.1177/1948550618774772

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). Springer.

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*(4), 257–267. https://doi.org/10.1016/J.CEDPSYCH.2011.05.002

Reid, K. (2005). The Causes, Views and Traits of School Absenteeism and Truancy [Article]. *Research in Education (Manchester)*, *74*(1), 59–82. https://doi.org/10.7227/RIE.74.6

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287–297. https://doi.org/10.1037//1040-3590.12.3.287

Renninger, K. A., & Bachrach, J. E. (2015). Studying Triggers for Interest and Engagement Using Observational Methods [Article]. *Educational Psychologist*, *50*(1), 58–69. https://doi.org/10.1080/00461520.2014.999920

Revelle, W. (2020). *psych: Procedures for personality and psychological research (Version 1.9.7)*. http://www.personality-project.org/r/psych/psych-manual.pdf

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. http://www.ncbi.nlm.nih.gov/pubmed/12405138

Rodrigo, M. M. T., Ryan S.J.d. Baker, & Rossi, L. (2013). Student Off-Task Behavior in Computer-Based Learning in the Philippines: Comparison to Prior Research in the USA. *Teachers College Record*, *115*(10).

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The Influence of Affective Teacher-Student Relationships on Students' School Engagement and Achievement: A Meta-Analytic Approach. *Review of Educational Research*. https://doi.org/10.3102/0034654311421793

Roos, A.-L., Bieg, M., Götz, T., Frenzel, A. C., Taxer, J., & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies*, *26*(2), 245–258. https://doi.org/10.1080/13598139.2015.1095078.

Rosário, P., Núñez, J. C., Vallejo, G., Cunha, J., Azevedo, R., Pereira, R., Nunes, A. R., Fuentes, S., & Moreira, T. (2016). Promoting Gypsy children school engagement: A story-tool project to enhance self-regulated learning. *Contemporary Educational Psychology*, *47*, 84–94. https://doi.org/10.1016/J.CEDPSYCH.2015.11.005.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Rosenholtz, S. J., & Rosenholtz, S. H. (1981). Classroom Organization and the Perception of Ability [Article]. *Sociology of Education*, *54*(2), 132–140. https://doi.org/10.2307/2112357

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom [Article]. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender Differences in the Effects of a Utility-Value Intervention to Help Parents Motivate Adolescents in Mathematics and Science [Article]. *Journal of Educational Psychology*, *107*(1), 195–206. https://doi.org/10.1037/a0036981

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: themes and variations*. Oxford University Press.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective Reading Instruction for Struggling Readers: The Role of Direct/Explicit Teaching. *Reading & Writing Quarterly*, *25*(2–3), 125–138. https://doi.org/10.1080/10573560802683523.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: basic psychological needs in motivation, development, and wellness*.

Sakata, N. (2022). Is learner-centred pedagogy associated with pupils' positive attitudes towards learning? The case of Tanzania [Article]. *Compare*, *ahead-of-p*(ahead-of-print), 1–19. https://doi.org/10.1080/03057925.2022.2036592

Salmela-Aro, K., Upadyaya, K., Cumsille, P., Lavonen, J., Avalos, B., & Eccles, J. S. (2021). Momentary task-values and expectations predict engagement in science among Finnish and Chilean secondary school students [Article]. *International Journal of Psychology*, *56*(3), 415–424. https://doi.org/10.1002/ijop.12719

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2007). *Assessment in special and inclusive education* (10th ed.). Houghton Mifflin Co.

Samuelson, B. L., & Freedman, S. W. (2010). Language policy, multilingual education, and power in Rwanda. *Language Policy*, *9*(3), 191–215. https://doi.org/10.1007/s10993-010-9170-7

Sandefur, J. (2018). Internationally comparable mathematics scores for fourteen african countries. *Economics of Education Review*, *62*, 267–286. https://doi.org/10.1016/J.ECONEDUREV.2017.12.003

Sanders, J. R., Fitzpatrick, J. L., & Worthen, B. R. (2010). *Program evaluation: alternative approaches and practical guidelines*. Longman.

Sayeda, Y., & Kanjeeb, A. (2013). Assessment in Sub-Saharan Africa: Challenges and Prospects. *Assessment in Education: Principles, Policy & Practice*, *20*(4), 373–384.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321. https://doi.org/10.1177/0734282911406653

Schmitz, B. (2006). Advantages of studying processes in educational research [Article]. *Learning and Instruction*, *16*(5), 433–449. https://doi.org/10.1016/j.learninstruc.2006.09.004

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is General Self-Efficacy a Universal Construct?: Psychometric Findings from 25 Countries. *European Journal of Psychological*

Assessment, *18*, 242–251. https://www.scienceopen.com/document?vid=f03a07db-3c96-437c-90b4-9ff335ffbd82

Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.

Schultzberg, M., & Muthén, B. (2018). Number of Subjects and Time Points Needed for Multilevel Time-Series Analysis: A Simulation Study of Dynamic Structural Equation Modeling [Article]. *Structural Equation Modeling*, *25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Schunk, D. H. (2013). *Learning theories : an educational perspective*. Pearson.

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2013). *Motivation in education: theory, research, and applications*. Pearson.

Schweisfurth, M. (2011). Learner-centred education in developing country contexts: From solution to problem? [Article]. *International Journal of Educational Development*, *31*(5), 419–426. https://doi.org/10.1016/j.ijedudev.2011.03.005

Schweisfurth, M. (2013). *Learner-centred education in international perspective: whose pedagogy for whose development?* Routledge.

Seidman, E., Kim, S., Raza, M., Ishihara, M., & Halpin, P. F. (2018). Assessment of pedagogical practices and processes in low and middle income countries: Findings from secondary school classrooms in Uganda [Article]. *Teaching and Teacher Education*, *71*, 283–296. https://doi.org/10.1016/j.tate.2017.12.017

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). Guilford Press.

Shapiro, E. S. (2011). *Academic skills problems workbook*. Guilford.

Shavelson, R. J., & Bolus, R. (1982). Self concept: The interplay of theory and methods. *Journal of Educational Psychology*, *74*(1), 3–17. https://doi.org/10.1037/0022-0663.74.1.3

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *6*(3), 7–1. http://journals.sagepub.com/doi/pdf/10.3102/00346543046003407

Shaver, P. R., Schwartz, J. C., Kirson, D., & O`Connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. http://www.ncbi.nlm.nih.gov/pubmed/3598857

Shernoff, D. J., Ruzek, E. A., & Sinha, S. (2017). The influence of the high school classroom environment on learning as mediated by student engagement. *School Psychology International*, *38*(2), 201–218. https://doi.org/10.1177/0143034316666413

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment [Article]. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual Stability in the Organization and Patterning of Behavior: Incorporating Psychological Situations Into the Idiographic Analysis of Personality [Article]. *Journal of Personality and Social Psychology*, *67*(4), 674–687. https://doi.org/10.1037/0022-3514.67.4.674

Simms, L. J. (2008). Classical and Modern Methods of Psychological Scale Construction. *Social and Personality Psychology Compass*, *2*(1), 414–433. https://doi.org/10.1111/j.1751-

9004.2007.00044.x

Simpson, J. (2014). *Baseline assessment of English language proficiency of school teachers in Rwanda.*

Sini, B., Muzzulini, B., Schmidt, S., & Tinti, C. (2018). School motivation: A comparison between Kenya and Italy. *The Journal of Educational Research*, *111*(6), 746–755. https://doi.org/10.1080/00220671.2018.1427035.

Skaalvik, E. M., & Rankin, R. J. (1995). A Test of the Internal/External Frame of Reference Model at Different Levels of Math and Verbal Self-Perception [Article]. *American Educational Research Journal*, *32*(1), 161–184. https://doi.org/10.3102/00028312032001161

Skinner, E. A. (1996). A Guide to Constructs of Control. *Journal of Personality and Social Psychology.*, *71*(3), 549–570. http://ezproxy-prd.bodleian.ox.ac.uk:2081/sp-3.27.2b/ovidweb.cgi?&S=LHKBFPPKKKDDEJPANCFKLELBKNNEAA00&Link+Set=S.sh.22.23.27.31%7C10%7Csl_10

Skinner, E. A. (2016). Engagement and disaffection as central to processes of motivational resilience development. In K. R. Wentzel & D. Miele (Eds.), *Handbook of motivation at school* (pp. 145–168). Lawrence Erlbaum Associates.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581. https://doi.org/10.1037/0022-0663.85.4.571

Skinner, E. A., Furrer, C. J., Marchand, G., & Kindermann, T. A. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765–781. https://doi.org/10.1037/a0012840

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement as an organizational construct in the dynamics of motivational development. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of Motivation at School* (pp. 223–245). Malwah, NJ: Erlbaum.

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In *Handbook of research on student engagement* (pp. 21–44).

Smit, K., de Brabander, C. J., & Martens, R. L. (2014). Student-centred and teacher-centred learning environment in pre-vocational secondary education: Psychological needs, and motivation [Article]. *Scandinavian Journal of Educational Research*, *58*(6), 695–712. https://doi.org/10.1080/00313831.2013.821090

Sommet, N., & Morselli, D. (2017). Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*(1), 203–218. https://doi.org/10.5334/irsp.90

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/JESP.1998.1373

Stallings, J. A. (1977). *Learning to Look: A Handbook on Classroom Observation and Teaching Models*. Wadsworth Publishing.

Stambach, A. (1994). "Here in Africa, We Teach; Students Listen": Lessons about Culture from Tanzania. *Journal of Curriculum and Supervision*, *9*(4).

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct Meaning in Multilevel Settings [Article]. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520.

https://doi.org/10.3102/1076998616646200

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement [Article]. *Learning and Individual Differences*, *19*(1), 80–90. https://doi.org/10.1016/j.lindif.2008.05.004

Stiefelhagen, R. (2002). Tracking focus of attention in meetings. *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference*, 273–280.

Stipek, D. J. (2002). Good instruction is motivating. In *Development of achievement motivation* (pp. 309–332). https://cepa.stanford.edu/content/good-instruction-motivating

Stocker, J., Abu-Hilal, M. M., Hermena, E., AlJassmi, M., & Barbato, M. (2021). Internal/external frame of reference model and dimensional comparison theory: a novel exploration of their applicability among Arab high school students [Article]. *Educational Psychology*, *41*(4), 483–501. https://doi.org/10.1080/01443410.2021.1887455

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the lavaan/semTools Packages. *Structural Equation Modeling*, *27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Swann, W. B., Chang-Schneider, C., & Mcclarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life [Article]. *American Psychologist*, *62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Tabulawa, R. (2013). *Teaching and learning in context: why pedagogical reforms fail in Sub-Saharan Africa*. CODESRIA.

Tafere, Y. (2014). *Education Aspirations and Barriers to Achievement for Young People in Ethiopia. An International Study of Childhood Poverty.* www.younglives.org.uk

Teasley, M. L. (2004). Absenteeism and Truancy: Risk, Protection, and Best Practice Implications for School Social Workers [Article]. *Children & Schools*, *26*(2), 117–128. https://doi.org/10.1093/cs/26.2.117

The R Core Team. (2019). *R: A Language and Environment for Statistical Computing. Version 3.6.2.* https://www.gnu.org/copyleft/gpl.html.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*(1), 144–151. https://doi.org/10.1037/0022-0663.92.1.144

Tobias Koch, Martin Schultze, Michael Eid, & Christian Geiser. (2014). A Longitudinal Multilevel CFA-MTMM Model for Interchangeable and Structurally Different Methods [Article]. *Frontiers in Psychology*, *5*, 311. https://doi.org/10.3389/fpsyg.2014.00311

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, *98*(2), 438–456. https://doi.org/10.1037/0022-0663.98.2.438

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*(3), 763–777. https://doi.org/10.1037/a0027470

Triandis, H. C., Malpass, R. S., & Davidson, A. R. (1971). Cross-Cultural Psychology. In *Review of Anthropology* (Vol. 7).

Tsai, Y.-M., Kunter, M., Ludtke, O., Ryan, R. M., & Trautwein, U. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. [Article]. *Journal of Educational Psychology*, *100*(2), 460. https://doi.org/10.1037/0022-0663.100.2.460

UNESCO. (2014). *Education for All Global Monitoring Report 2013/14, Teaching and Learning: Achieving Quality for All*.

Upadyaya, K., Cumsille, P., Avalos, B., Araneda, S., Lavonen, J., & Salmela-Aro, K. (2021). Patterns of situational engagement and task values in science lessons [Article]. *The Journal of Educational Research*, *114*(4), 394–403. https://doi.org/10.1080/00220671.2021.1955651

Urdan, T. C., & Maehr, M. L. (1995). Beyond a Two-Goal Theory of Motivation and Achievement: A Case for Social Goals. *Review of Educational Research*, *65*(3), 213–243. https://doi.org/10.3102/00346543065003213

Urdan, T. C., & Turner, J. C. (2005). Competence Motivation in the Classroom. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 297–317). Guilford Press.

Usher, E. L. (2015). Personal Capability Beliefs. In L. Corno & E. M. Anderman (Eds.), *Handbook of Educational Psychology* (pp. 146–159). Routledge Handbooks Online. https://doi.org/10.4324/9781315688244.ch11

Usher, E. L., & Pajares, M. F. (2009). Sources of self-efficacy in mathematics: A validation study [Article]. *Contemporary Educational Psychology*, *34*(1), 89–101. https://doi.org/10.1016/j.cedpsych.2008.09.002

Vallerand, R. J. (1997). Toward A Hierarchical Model of Intrinsic and Extrinsic Motivation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 29, pp. 271–360). Academic Press. https://doi.org/10.1016/S0065-2601(08)60019-2

van Buuren, S., & Groothuis-Oudshoorn, C. G. M. (2011). mice: Multivariate Imputation by Chained Equations in R [Article]. *Journal of Statistical Software*, *45*(3), urn:issn:1548-7660. https://doi.org/10.18637/jss.v045.i03

van de Kuilen, H. S., Altinyelken, H. K., Voogt, J. M., & Nzabalirwa, W. (2019). Policy adoption of learner-centred pedagogy in Rwanda: A case study of its rationale and transfer mechanisms. *International Journal of Educational Development*, *67*, 64–72. https://doi.org/10.1016/j.ijedudev.2019.03.004

van de Kuilen, H. S., Altinyelken, H. K., Voogt, J. M., & Nzabalirwa, W. (2020). Recontextualization of learner-centred pedagogy in Rwanda: A comparative analysis of primary and secondary schools [Article]. *Compare*, 1–18. https://doi.org/10.1080/03057925.2020.1847044

van de Schoot, R., Kaplan, D., Denissen, J. J. A., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*, *85*(3), 842–860. https://doi.org/10.1111/cdev.12169

van der Westhuizen, L., Arens, A. K., Greiff, S., Fischbach, A., & Niepel, C. (2022). The generalized internal/external frame of reference model with academic self-concepts, interests, and anxieties in students from different language backgrounds [Article]. *Contemporary Educational Psychology*, *68*, 102037. https://doi.org/10.1016/j.cedpsych.2021.102037

Vansteenkiste, M., Sierens, E., Goossens, L., Soenens, B., Dochy, F., Mouratidis, A., Aelterman, N., Haerens, L., & Beyers, W. (2012). Identifying configurations of perceived teacher autonomy support and structure: Associations with self-regulated learning, motivation and problem behavior. *Learning and Instruction*, *22*(6), 431–439. https://doi.org/10.1016/J.LEARNINSTRUC.2012.04.002

Vasalampi, K., Muotka, J., Malmberg, L.-E., Aunola, K., & Lerkkanen, M.-K. (2020). Intra-individual dynamics of lesson-specific engagement: Lagged and cross-lagged effects from one lesson to the next [Article]. *The British Journal of Educational Psychology*, e12404. https://doi.org/10.1111/bjep.12404

Vasalampi, K., Muotka, J., Pöysä, S., Lerkkanen, M.-K., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Assessment of students' situation-specific classroom engagement by an InSitu Instrument [Article]. *Learning and Individual Differences*, *52*, 46–52. https://doi.org/10.1016/j.lindif.2016.10.009

Venäläinen, R. (2008). *What do we know about instructional time use in Mali? Assessing the suitability of the classroom observation snapshot instrument for use in developing countries*. http://documents.worldbank.org/curated/en/172371468158067336/What-do-we-know-about-instructional-time-use-in-Mali-Assessing-the-suitability-of-the-classroom-observation-snapshot-instrument-for-use-in-developing-countries

Vogt, W. P. (2011). SAGE quantitative research methods. In *Quantitative research methods*. SAGE.

Volpe, R. J., DiPerna, J., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, *34*(4), 454–474. https://scholarworks.umass.edu/education_faculty_pubs/197

Walls, T. A. (2013). Intensive Longitudinal Data. In *The Oxford handbook of quantitative methods in psychology. Vol. 2, Statistical analysis* (pp. 432–440). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199934898.013.0020.

Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data*. Oxford University Press.

Walton, R., Zraly, M., & Mugengana, J. P. (2015). Values and Validity: Navigating Messiness in a Community-Based Research Project in Rwanda. *Technical Communication Quarterly*, *24*(1), 45–69. https://doi.org/10.1080/10572252.2015.975962.

Wang, M.-T., Guo, J., & Degol, J. L. (2019). The Role of Sociocultural Factors in Student Achievement Motivation: A Cross-Cultural Review [Article]. *Adolescent Research Review*. https://doi.org/10.1007/s40894-019-00124-y

Wang, X., & Cheng, Z. (2020). Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations [Article]. *Chest*, *158*(1), S65–S71. https://doi.org/10.1016/j.chest.2020.03.012

Warrington, M., & Kiragu, S. (2012). "It makes more sense to educate a boy": Girls 'against the odds' in Kajiado, Kenya. *International Journal of Educational Development*, *32*(2), 301–309. https://doi.org/10.1016/j.ijedudev.2011.05.004

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, *79*(1), 1–28. https://doi.org/10.1348/000709908X380772

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology*, *90*(2), 202–209. https://doi.org/10.1037/0022-0663.90.2.202

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In Rick H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). The Guilford Pressb.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*(1), 1–35. https://doi.org/10.1016/J.DR.2009.12.001

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/CEPS.1999.1015

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S. D., Roeser, R. W., & Schiefele, U. (2015). Development of Achievement Motivation and Engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of Child Psychology and Developmental Science. Volume 3: Socioemotional Processes.* (pp. 657–700). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0315

Wigfield, A., Eccles, J. S., & Möller, J. (2020). How Dimensional Comparisons Help to Understand Linkages Between Expectancies, Values, Performance, and Choice [Article]. *Educational Psychology Review*, *32*(3), 657–680. https://doi.org/10.1007/s10648-020-09524-2

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In W. (Series E. . Damon & N. (Volume E. . Eisenberg (Eds.), *Handbook of child psychology, 6th Ed. Vol.3. Social, emotional, and personality development* (pp. 933–1002). John Wiley & Sons, Ltd.

Wigfield, A., & Gladstone, J. R. (2019). What Does Expectancy-value Theory Have to Say about Motivation and Achievement in Times of Change and Uncertainty? In E. N. Gonida & M. S. Lemos (Eds.), *Motivation in Education at a Time of Global Change* (pp. 15–32). https://doi.org/10.1108/S0749-742320190000020002

Wigfield, A., Tonks, S. M., & Klauda, S. T. (2016). Expectancy-value theory. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 55–74). Routledge.

Wiley, D. E., & Harnischfeger, A. (1974). Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles [Article]. *Educational Researcher*, *3*(4), 7–12. https://doi.org/10.2307/1175988

Wilkins, J. L. M. (2004). Mathematics and Science Self-Concept: An International Investigation [Article]. *The Journal of Experimental Education*, *72*(4), 331–346. https://doi.org/10.3200/JEXE.72.4.331-346

Williams, T. P. (2015). *The things they learned: aspiration, uncertainty, and schooling in rural Rwanda. A thesis submitted for the degree of Doctor of Philosophy.* https://canvas.harvard.edu/courses/6972/files/1285368/download?verifier=PRpv9ABlvqKo5kUoc2HJViPUVj5yZnRk31dTuiWO&wrap=1

Williams, T. P. (2019). The Things They Learned: Aspiration, Uncertainty, and Schooling in Rwanda's Developmental State [Article]. *The Journal of Development Studies*, *55*(4), 645–660. https://doi.org/10.1080/00220388.2018.1453602

Williams, T. P., Abbott, P., & Mupenzi, A. (2015). 'Education at our school is not free': the hidden costs of fee-free schooling in Rwanda. *Compare: A Journal of Comparative and International*

Education, *45*(6), 931–952. https://doi.org/10.1080/03057925.2014.938611.

Willis, G. B. (2005). *Cognitive interviewing: a tool for improving questionnaire design*. SAGE.

Winograd, P., & Chou, H. V. (1988). Direct instruction of reading comprehension strategies: The nature of teacher explanation. In C. Weinstein, E. Goetz, & P. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 121–139). Academic Press, Inc.

Wolff, F., Lüdtke, O., Helm, F., & Möller, J. (2021). Integrating the big-fish-little-pond effect, the basking-in-reflected-glory effect, and the internal/external frame of reference model predicting students' individual and collective academic self-concepts [Article]. *Contemporary Educational Psychology*, *65*, 101952. https://doi.org/10.1016/j.cedpsych.2021.101952

Wolff, F., Nagy, N., Helm, F., & Möller, J. (2018). Testing the internal/external frame of reference model of academic achievement and academic self-concept with open self-concept reports [Article]. *Learning and Instruction*, *55*, 58–66. https://doi.org/10.1016/j.learninstruc.2017.09.006

Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2004). Cognitive Pretesting and the Developmental Validity of Child Self-Report Instruments: Theory and Applications [Article]. *Research on Social Work Practice*, *14*(3), 191–200. https://doi.org/10.1177/1049731503257882

World Bank. (2015). *Conducting Classroom Observations. Analyzing Classroom Dynamics and Instructional Time. Using the Stallings "Classroom Snapshot" Observation System*. http://documents.worldbank.org/curated/en/790221467997639302/Conducting-classroom-observations-analyzing-classrooms-dynamics-and-instructional-time-using-the-Stallings-classroom-snapshot-observation-system-user-guide

Wragg, E. C. (2012). *An introduction to classroom observation* (Classic ed). Routledge.

Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes [Article]. *Psychometrika*, *81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Wu, X., Anderson, R. C., Nguyen-Jahiel, K., & Miller, B. (2013). Enhancing motivation and engagement through collaborative discussion [Article]. *Journal of Educational Psychology*, *105*(3), 622–632. https://doi.org/10.1037/a0032792

Xu, M. K., Marsh, H. W., Hau, K.-T., Ho, I. T., Morin, A. J. S., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction [Article]. *Journal of Educational Psychology*, *105*(2), 489–503. https://doi.org/10.1037/a0031333

Yeager, D. S., & Walton, G. M. (2011). Social-Psychological Interventions in Education: They're Not Magic [Article]. *Review of Educational Research*, *81*(2), 267–301. https://doi.org/10.3102/0034654311405999

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview [Article]. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

Zusho, A., & Clayton, K. E. (2011). Culturalizing Achievement Goal Theory and Research. *Educational Psychologist*, *46*(4), 239–260. https://doi.org/10.1080/00461520.2011.614526.

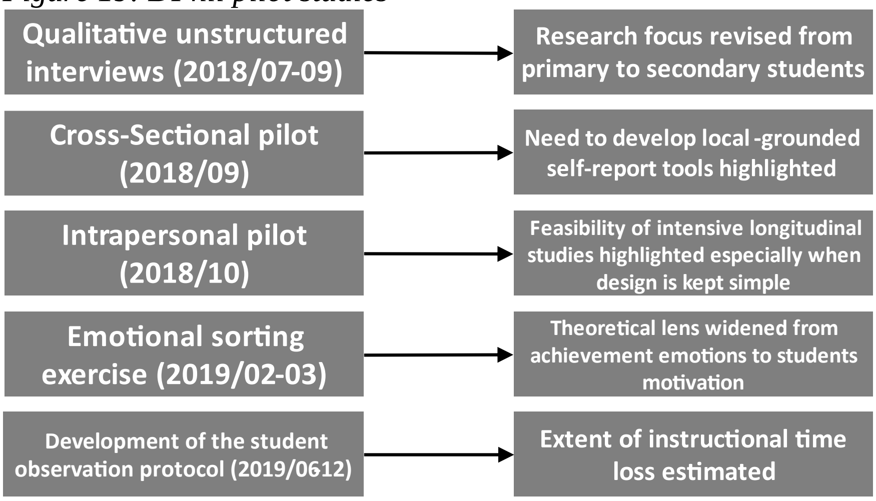
Annex 1: Early Pilot Studies

As per the initial research design of this DPhil project, the focus was on applying Pekrun's control-value theory to the learning amongst Rwandan primary students. According to the control-value theory, achievement emotions are assumed to drive amongst others a person's level of engagement exhibited during an achievement situation (i.e., cognitive and behavioural processes). These emotions in turn are hypothesized to be the immediate product of internal subjective evaluations ("appraisals") people form about themselves within achievement situations. The appraisals concern one's own perceived competencies to perform those learning activities and to attain those learning goals ("control appraisals) that constitute the learning challenge as well as the perceived interestingness or personal importance those challenges carry for oneself ("value appraisals) (Pekrun, 2006; Pekrun et al., 2007; Pekrun & Perry, 2014). Moreover, initially it was planned to administer a questionnaire twice during lessons (i.e., pre and post questionnaire) in order to be able to estimate the mediational role of achievement emotions within motivational processes. Based on a series of pilot studies, however, this initial design was eventually revised in the way outlined in section 3. In order to document how the overall research design and focus of this DPhil project evolved over time, these initial pilot studies are summarized below. For each of these study, overall method, key findings, and key take-away points concerning the DPhil project are outlined. Figure A1 below presents schematic overview of the various pilot phases and main take-away points associated with each.

Annex 1.1: Qualitative Studies, 2018

Preparatory fieldwork for this DPhil project started in July 2018. At the beginning, it mainly comprised of classroom observations of primary classes, the initial sample focus of the project, and repeated semi-structured interviews with students sampled out of those classes. The former helped to better understand how typical lessons at public schools in Rwanda look like, something that was useful especially during the development of the classroom observation protocol (see annex 2). To carry out these semi-structured interviews, nine upper primary students (six females and three males; aged around 10) from two different public schools, one from within Kigali and the other just outside Kigali, were recruited. These students were supposed to be visited repeatedly over the next weeks. It was planned on each occasion to discuss with these children two main points. The first point would have been about what about school on that day made them feel good (i.e., positive emotions) or bad (i.e., negative emotions). The second point should then have been about the reasons for feeling good or bad. That way, it was hoped to explore how primary students experience and express achievement emotions (point one) as well as motivational appraisals such as control and value evaluations (point two).

Figure 19: DPhil pilot studies



Especially with hindsight, the semi-structured interviews with primary students were not considered successful. According to their answers, students only felt very positive about their daily school experience all the time without having any bad experience. In a similar vein, they described everything they learn at school as very useful and important. According to them, they always felt capable to complete all the learning tasks because they “always work so hard” (female, primary-two student). Of course, it is not to say that Rwandan students could not have largely favourable experience at their public schools. It was just that the extent of positive experience was considered exaggerated. One reason for this overreporting may have been that primary students perceived these interviews to be “interrogations” (McWilliam et al., 2009, p. 70) and thus wanted to provide those answers they deemed to be acceptable in the eyes of this researchers (i.e., social desirability bias). Of course, all participants were repeatedly reminded that this DPhil research was solely about exploring students’ actual experiences at school rather than getting any socially desirable answers about their study habits. However, they might have been still too young to fully understand this. Finally, due to Rwanda’s double-shift system, which is only implemented at the primary level, students have classes alternately in the morning and the afternoon. Classroom observations revealed that especially in the afternoon many students, and their teachers, only attend school very irregular, something that would have rendered the organization of an intrapersonal study very difficult. Hence, it was decided to shift the focus of this DPhil from primary to secondary students.

Annex 1.2: Cross-Sectional Pilot, 2018

A second preparatory study was a quantitative cross-sectional survey. It had the following two goals. First, it helped to explore the general willingness to partake in motivational research amongst Rwandan public-school students. Second, it provided a glimpse on general response patterns within questions on motivational appraisals, achievement emotions, and student engagement. The survey was carried out at one of Kigali's public schools targeting students from primary four through secondary two towards the end of the end of the 2018 school year (i.e., September). The school was randomly selected out of Kigali's school register. Secondary-3 students were not included because at that time they were in the final revision period before their end-of-school exams. Across all target classes, 770 students were enrolled.

To recruit students, all target classes were visited prior to data collection. On those visits, the purpose of the study was introduced, the data-collection procedure outlined, and rights of participants explained. Participation was voluntary and was not tied to any form of benefit. On the day of data collection, target classes were visited one by one. Surveys were handed out by this DPhil student as well as his research assistant, a female Rwanda. 201 students that had previously submitted valid consent forms completed the questionnaire (i.e., overall participation rate of 26.1 percent). Survey participation was not tied to any material benefits. In general, primary students appeared to be more likely to partake in the study (the dates of the survey clashed with some mock exams students in primary six had to attend, which is the main reason why their participation rates appear to be lower compared to especially P4 and P5). After the survey, some informal interviews with secondary students were carried out. Students recommended the provision of incentives (i.e., lottery games) to motivate student participation. This recommendation has been incorporated into the design of the main study.

Data collection targeted the subjects of Math and Kinyarwanda, Rwanda's lingua franca. Motivational appraisals, achievement emotions, and behavioural engagement were measured through eight scales altogether that all relied on students' self-reports. Motivational appraisals consisted of two scales. The first one was about control appraisals (three items; e.g., “Work in [SUBJECT] is easy for me”) and was adapted from Arens et al. (2013). The second appraisal scale was about value evaluations (three items; e.g., “I am interested in the things I learn in [SUBJECT]”) and was taken from OECD (2013a). Four scales were about achievement emotions and measured anxiety (three items; e.g., “I feel nervous in the [SUBJECT] class.”), boredom (four items; e.g., “I feel like going out during the [SUBJECT] class because it is so boring”), enjoyment (two items; e.g., “I feel good when I am in the [SUBJECT] class listening to the teacher talk.”), and pride (two items; e.g., “I take pride in being able to keep up with the material in the [SUBJECT] class.”). Emotion scales were adjusted from Peixoto et al. (2015) and Pekrun et al. (2005). Student engagement included both efforts spent in class as well as during homework. The former scale (6 items; e.g., When I’m in [SUBJECT], I participate in class discussions.) was adjusted from Skinner et al. (2008) whereas the latter one (6 items; e.g., “I do my [SUBJECT] homework to the best of my ability.”) and was adjusted from Trautwein et al. (2006). The questionnaire was administered in Kinyarwanda using only a paper version. Both engagement scales contained both positively (e.g., In [SUBJECT] class, I work as hard as I can.) and negatively worded items (e.g., “When I’m in [SUBJECT], my mind wanders.”). All answer options ranged from one (most of the time) to four (never).

Table 55 and 54 on the next page provide summary statistics on students’ beliefs, achievement emotions, (dis)engagement within Kinyarwanda and maths. Standard deviations suggest that there is some variation in the answers provided. Especially younger children exhibit larger levels of variation in their replies than older children. Also, basic correlational analysis (not presented here) reveals that at the trait level control and value beliefs generally predict achievement emotions, which in turn predict engagement. Both tables also exhibit basic psychometric properties (i.e., skewness and kurtosis) of the different scales. Skewness generally deviated substantially from 0 and kurtosis is frequently way below or above three (Vogt, 2011). It appears that with the sole exception of primary six in Kinyarwanda and primary five in Math, scales generally do not exhibit excess kurtosis. Moreover, it appears that sales are less skewed within the secondary compared to the primary data. Nonetheless, all scales employed exhibit lower reliabilities amongst others expressed in form of McDonald's omega (hierarchical). Within Kinyarwanda, coefficients across the eight scales are around 0.51 whereas in Math is even only around 0.31. These unreliabilities may also explain lower factor correlations. Confirmatory factor analysis suggests that latent factors representing appraisals, emotions, and engagement are hardly correlated within Kinyarwanda. In Math, however, especially control appraisals were predicting negatively anxiety and boredom as well as positively predicting enjoyment as well as pride. Moreover, negative emotions were predicting negatively and positive emotions predicting positively especially classroom-based engagement. However, respective correlations were always weak. This all may suggest that self-report instruments from the wider education literature perform not great in measuring motivational beliefs, emotions, and engagement. For this reason, it was decided not to rely on typical measures from the motivational literature but to develop self-report instruments from scratch through a bottom-up approach in the way outlined in section 3. Besides this important learning insight, however, a key takeaway point was that the implementation of a cross-sectional survey to measure students' traits is indeed feasible also within the context of this DPhil project.

Table 54: Students' beliefs, emotions, and engagement in Kinyarwanda (pilot phase)

Scale	Primary 4				Primary 5				Primary 6				Secondary 1				Secondary 2			
	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR
Control Beliefs	1.48	0.71	1.82	6.04	1.64	0.56	0.85	2.98	1.59	0.62	1.08	3.72	1.27	0.34	1.09	3.59	1.55	0.59	1.33	4.59
Value Beliefs	1.73	0.75	0.79	2.54	1.24	0.36	1.76	5.66	1.26	0.60	2.95	12.02	1.32	0.45	1.34	3.66	1.36	0.45	1.20	3.23
Anxiety	2.44	0.72	-0.02	2.52	2.87	0.75	-0.17	2.03	2.56	0.89	0.17	1.93	2.93	0.76	-0.50	2.65	2.60	0.81	-0.41	2.51
Boredom	2.63	1.00	-0.06	1.79	3.34	0.68	-0.91	3.23	3.48	0.80	-1.62	5.01	3.45	0.70	-1.61	5.49	3.32	0.67	-0.72	2.16
Enjoyment	1.46	0.63	0.97	2.46	1.30	0.44	1.20	3.23	1.29	0.49	2.07	7.36	1.52	0.51	0.48	2.05	1.76	0.89	1.30	3.90
Pride	1.62	0.66	1.47	5.74	1.37	0.57	1.26	3.26	1.23	0.53	2.33	7.33	1.27	0.44	1.44	3.9	1.18	0.35	1.63	4.07
Class Engagement	1.95	0.68	0.53	2.70	2.09	0.46	0.41	3.81	2.07	0.41	1.38	8.15	2.08	0.33	-0.17	3.02	2.02	0.25	0.36	3.64
Homework Engagement	2.01	0.59	0.20	2.59	2.03	0.42	0.34	3.79	1.99	0.43	0.26	4.43	1.97	0.35	0.29	2.42	2.20	0.52	0.45	3.02
N	55				53				24				42				25			

Note. 'SE' refers to standard errors, 'SKE' refers to skewness, and 'KUR' refers to kurtosis. 'N' refers to sample size.

Table 55: Students' beliefs, emotions, and engagement in mathematics (pilot phase)

Scale	Primary 4				Primary 5				Primary 6				Secondary 1				Secondary 2			
	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR	Mean	SE	SKE	KUR
Control Beliefs	1.41	0.57	1.68	5.48	1.49	0.64	1.34	4.14	1.84	0.70	1.46	5.55	2.42	0.78	-0.28	2.43	2.19	0.66	-0.03	2.31
Value Beliefs	1.47	0.54	0.87	2.84	1.23	0.51	2.29	7.31	1.24	0.38	1.09	2.55	1.31	0.46	1.35	3.72	1.30	0.60	2.59	8.74
Anxiety	2.48	1.01	0.25	1.88	2.88	0.85	-0.38	2.30	2.24	0.87	0.04	1.98	2.58	0.82	-0.37	2.36	2.22	0.67	-0.02	2.24
Boredom	2.69	1.04	-0.25	1.78	3.46	0.68	-1.02	2.96	3.37	0.82	-1.52	4.58	2.82	0.82	-0.14	2.12	3.02	0.85	-0.61	2.65
Enjoyment	1.39	0.53	0.83	2.11	1.18	0.44	2.32	6.91	1.48	0.62	0.93	2.69	1.96	0.86	0.35	1.99	1.60	0.74	1.65	5.78
Pride	1.64	0.81	1.04	3.14	1.27	0.64	2.56	9.20	1.38	0.77	2.26	7.31	1.89	0.79	0.55	2.24	1.32	0.52	1.91	6.20
Class Engagement	1.90	0.61	0.69	4.62	2.11	0.37	-0.10	3.71	2.11	0.37	0.97	3.68	2.10	0.31	0.33	3.54	2.05	0.38	0.76	3.11
Homework Engagement	1.97	0.64	0.67	4.35	1.95	0.41	1.33	8.26	2.00	0.30	-0.04	1.91	2.06	0.42	0.05	2.26	2.29	0.50	0.36	2.34
N	55				53				24				42				25			

Note. 'SE' refers to standard errors, 'SKE' refers to skewness, and 'KUR' refers to kurtosis. 'N' refers to sample size.

Annex 1.3: Intrapersonal Pilot, 2018

A cross-sectional survey is slightly less challenging from a logistical point of view than an intensive-longitudinal school study. The latter implies repeated data collection over time involving the same sample. Hence, respondents need to be tracked across different measurement occasions. Moreover, intensive-longitudinal studies require each round of data collection to be taken place ideally close in time and location to the lesson they target. Especially when more than one school class is involved, respective studies also require careful coordination of data collection across different classrooms, and lesson periods. The latter point was considered particularly challenging within the Rwandan public-school context. The formal and informal classroom observations revealed that teachers are often not on time both in terms of the starting and finishing lessons. At times they were also found not to attend to a class at all. This rendered schooling less predictable than initially anticipated. In order to gain first practical experience in the implementation of an intrapersonal study within the Rwandan public-school system, it was therefore decided to carry out a small pilot prior to the main research phase. This intrapersonal pilot was carried out in October 2018, the last month of teaching of the corresponding school year.

Table 56: Summary statistics of lesson-specific data on English (pilot phase)

Scale	Primary 4				Primary 5				Primary 6			
	Mean	SE	SKE	KU	Mean	SE	SKE	KU	Mean	SE	SKE	KU
Control	3.70	0.47	-0.85	1.72	3.76	0.56	-2.25	6.93	3.19	0.81	-0.99	3.83
Value	4.00	0.00	NA	NA	3.95	0.22	-4.13	18.1	3.78	0.42	-1.38	2.90
Boredom	2.30	0.95	-0.38	1.65	2.15	0.99	0.03	1.64	1.81	0.91	1.06	3.44
Enjoyment	3.67	0.84	-2.46	7.70	3.60	0.94	-2.24	6.52	2.78	1.36	-0.41	1.36
Disengage.	2.47	1.14	0.01	1.64	1.85	1.04	0.59	1.78	1.41	0.80	2.14	6.95
Engagement	3.17	1.09	-0.83	2.20	3.60	0.68	-1.40	3.61	3.19	0.86	-0.66	2.42
Progress	3.37	0.81	-0.75	1.98	3.60	0.75	-1.48	3.48	3.24	0.95	-0.50	1.32
N	30				20				37			

Note. 'SE' refers to standard errors, 'SKE' refers to skewness, and 'KUR' refers to kurtosis. 'N' refers to sample size.

Data collection targeted a P4 class, a P5 class, and a P6 class at one of the public schools in a district outside Kigali. Per class targeted, 10 participants were randomly selected out of the corresponding student register. Before data collection, all target classes were visited in order to explain the procedures associated with this small-scale study to all students. In addition, all target students randomly selected were invited to participate in the study. In exchange they were all eligible to partake in a lottery game that was implemented after pilot completion. To partake in the study, consent forms were handed out to them, which they had to return in the following days. Informal feedback by students as well as the school headmaster suggested that especially parents' lack of interest (and potentially ability) to go through a research consent form resulted in children often being able to return consent forms only after some days. Hence, it was only through repeated follow-up with students, and with their parents, that all consent forms, duly signed, could be collected.

Table 57: Summary statistics of lesson-specific data (P6 only) in Kinyarwanda (pilot phase)

Scale	Primary 6			
	Mean	SE	SKEW	KUR
Control	3.56	0.56	-0.76	2.5
Value	3.76	0.43	-1.25	2.56
Boredom	1.61	1.06	1.48	3.64
Enjoyment	2.56	1.46	-0.09	1.08
Disengage.	1.5	1.08	1.83	4.52
Engagement	3.52	0.67	-1.02	2.87
Progress	3.15	0.95	-0.3	1.22
# data points		34		
# lessons		4		

Note. 'SE' refers to standard errors, 'SKE' refers to skewness, and 'KUR' refers to kurtosis. 'N' refers to sample size.

Once all forms were collected, a briefing session was organized for each target class. On that occasion, each target student was handed out a unique identification sheet, a sheet of paper with his/ her name written on it both sides. They were all instructed to have the sheets on their desks all the time. That way, it was possible to identify them at any point during the lessons. This worked as long as students did not lose their identification sheets, something that increasingly happened towards the end of the pilot. A general observation across all pilot phases was that especially in the afternoon both students and teachers are less likely to attend lessons (on time). It was therefore decided to carry out the intrapersonal pilot only in the morning. Since all primary classes are taught alternately in the morning or afternoon, all target students were therefore subject to only three days of data collection instead of 5. To maximise data points, lesson-specific data collection targeted not only Kinyarwanda and Math but all six school subjects.

Table 58: Summary statistics of lesson-specific data in mathematics (pilot phase)

	Primary 4				Primary 5				Primary 6			
	Mean	SE	SKEW	KUR	Mean	SE	SKEW	KUR	Mean	SE	SKEW	KUR
Control	3.98	0.15	-6.4	42.02	3.64	0.63	-1.5	4.1	3.35	0.51	0.27	1.9
Value	3.9	0.31	-2.63	7.91	3.84	0.37	-1.88	4.52	3.83	0.38	-1.74	4.03
Boredom	2.27	1.08	-0.04	1.55	2.32	1.16	0.02	1.53	1.47	0.8	1.81	5.73
Enjoyment	3.74	0.83	-2.97	10	3.32	1.38	-2.03	6.28	2.67	1.37	-0.25	1.24
Disengage.	2.46	1.11	-0.08	1.67	2.32	1.25	0.08	1.38	1.22	0.58	2.97	12.07
Engagement	3.15	1.09	-0.79	2.11	3.32	1.16	-1.31	3.01	3.25	0.8	-0.66	2.47
Progress	3.54	0.73	-1.23	2.98	3.68	0.75	-1.88	4.52	3.33	0.98	-0.9	2.13
# data points	50				19				64			
# lessons	5				2				7			

Note. ‘SE’ refers to standard errors, ‘SKE’ refers to skewness, and ‘KUR’ refers to kurtosis. ‘N’ refers to sample size.

Lesson-specific data collection implied students filling a paper-based questionnaire at the beginning and another one upon completion of a lesson targeted. The intrapersonal questionnaire was about students’ control and value appraisals, achievement emotions, and engagement; each construct was measured through one item only (e.g. Bieg et al., 2014; Götz et al., 2010, 2016). The item on control appraisal was adjusted from the scale on control beliefs mentioned above. The item on value appraisal was adapted from Goetz et al. (2010). The item on state boredom was adjusted from Pekrun (2005) whereas the item on enjoyment was taken from Patall et al. (2016). Behavioural engagement was measured through two items (i.e., disengagement and engagement) adjusted from Skinner et al. (2008). In addition, a single-item measure on perceived learning progress, adapted from Shernoff et al. (2017, p. 207) was included. Tables 56 to 60 presents respective item-specific summary statistics. Note, for some subjects more than 30 data points were collected since some students provided self-reports for some subjects more than once.

The statistical findings of the study were only of secondary interest. Amongst others it was expected that additional work would be required to improve the psychometric properties especially of the self-report instruments. Nonetheless, a variance-component analysis revealed that within-person variance is substantial across all appraisal, emotion, and engagement items. Hence, intrapersonal studies on motivation are warranted within the African context, too. Also, even though associations were not strong, multilevel modelling suggests that control appraisal predicts behavioural engagement, and value appraisal predicts emotional states (i.e., enjoyment). Of special interest for the main research phase, however, was that the pilot clearly demonstrated the feasibility of an intrapersonal study at Rwandan schools. However, it also showed that any intrapersonal design is best kept as simple as possible. For example, students are willing to fill lesson-specific questionnaires but administering a questionnaire twice per lesson would take up too much of a teacher's time. Two questionnaires per lesson may also have increased the risk of growing survey fatigue amongst study participants. It was therefore decided to reduce intrapersonal data collection to just one questionnaire per lesson. Also, identification sheets may not necessarily a viable option to identify students on an ongoing basis. Primary children may keep and use them for some days. But the longer a study runs the more likely that they lose them. Moving forward one would have to replace those sheets lost, something that in turn would add another layer of necessary organization. Instead, it would be the easiest if questionnaires will be given out to every student in class who wants one. Questionnaires of those students, which, for example, have not provided consent forms previously, are best filtered out upon completion of data collection. Merging data from different sources based on students' personal information such as name, age, gender, and class worked well and was considered to be a practical solution for the main study as well.

Table 59: Summary statistics of lesson-specific data on SET (pilot phase)

	Primary 4				Primary 5			
	Mean	SE	SKEW	KUR	Mean	SE	SKEW	KUR
Control	3.71	0.52	-1.51	4.36	3.9	0.31	-2.67	8.11
Value	0.87	0.41	-3.25	13.14	3.85	0.37	-1.96	4.84
Boredom	2.33	0.87	-0.45	1.91	2.05	1.15	0.33	1.50
Enjoyment	3.67	0.76	-2.53	8.61	3.56	0.92	-1.79	4.75
Disengage.	2.35	1.05	0.07	1.80	1.85	0.93	0.30	1.28
Engagement	3.27	0.96	-1.10	3.08	3.25	1.07	-1.04	2.64
Progress	3.38	0.90	-1.02	2.61	3.60	0.82	-1.50	3.25
# data points	40				20			
# lessons	4				2			

Note. 'SE' refers to standard errors, 'SKE' refers to skewness, and 'KUR' refers to kurtosis. 'N' refers to sample size.

Annex 1.4: Emotion Pilot Studies, 2019

As per the initial research design, a key hypothesis to be tested was that achievement emotions perform a mediational role within students' motivational processes at school. According to the control-value theory (e.g. Pekrun & Perry, 2014), at least 17 different types of discrete emotions can be distinguished. It is beyond the possibilities of this research project to control for all 17 discrete states within the intrapersonal study of on the mediational role of emotions. Due to pragmatic reasons, it was decided to focus on a smaller number of emotions. The previous cross-sectional and intrapersonal pilot focussed amongst others on boredom and enjoyment whose empirical operationalization through self-reports was derived from the Western-borne motivational literature. However, especially as the interpersonal pilot study revealed these scales seem not to resonate well with the Rwandan context. As a consequence, it was decided to develop motivational self-report scales (i.e., motivational beliefs and appraisals) bottom up through an indicative scale-development process. Moreover, it was intended to develop achievement-emotion scales that are grounded in the local context and culture following a prototype-research process.

Table 60: Summary Statistics of Lesson-Specific Data on Social Studies

	Primary 4				Primary 5				Primary 6			
	Mean	SE	SKEW	KUR	Mean	SE	SKEW	KUR	Mean	SE	SKEW	KUR
Control	3.75	0.55	-2.07	6.29	3.60	0.68	-1.40	3.61	3.20	0.73	-0.61	3.06
Value	3.90	0.31	-2.67	8.11	3.85	0.37	-1.96	4.84	3.93	0.26	-3.25	11.58
Boredom	2.30	1.03	-0.33	1.52	2.05	1.00	0.23	1.66	1.54	0.84	1.62	4.92
Enjoyment	3.85	0.67	-4.13	18.05	3.75	0.72	-3.14	12.18	2.89	1.30	-0.58	1.58
Disengage.	2.45	1.50	-0.52	2.27	1.75	0.97	0.52	1.33	1.62	1.15	1.46	3.31
Engagement	3.00	1.29	-0.64	1.65	3.25	1.25	-1.15	2.47	3.19	0.93	-1.08	3.36
Progress	3.65	0.67	-1.64	4.26	3.35	0.93	-0.75	1.63	2.95	0.99	0.11	1.05
# data points		20				20				55		

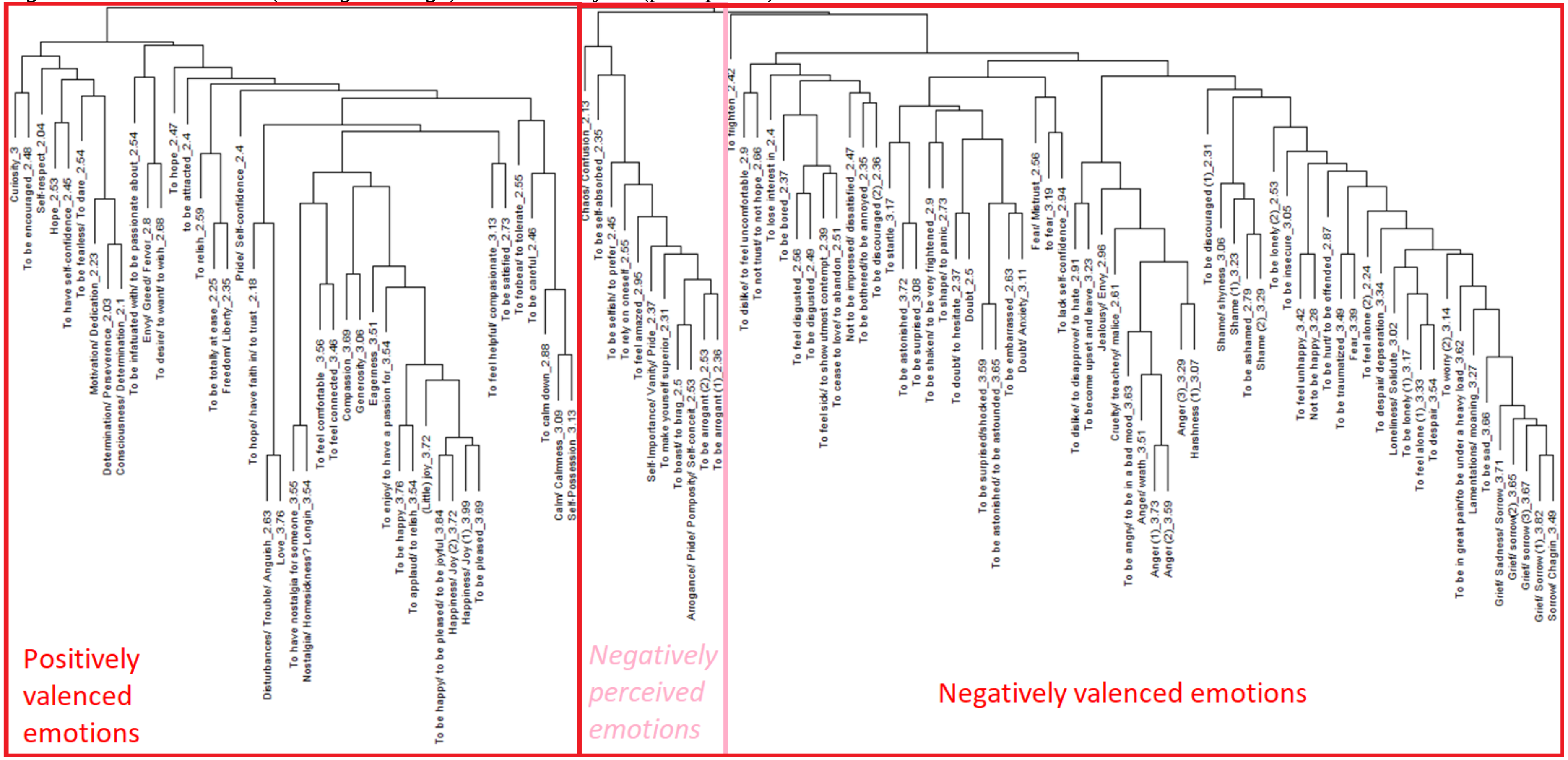
Note. ‘SE’ refers to standard errors, ‘SKE’ refers to skewness, and ‘KUR’ refers to kurtosis. ‘N’ refers to sample size.

Prototype researchers study the ways in which humans develop perceptions of the world surrounding them through forming hierarchical representations of categorical systems of objects and events they experience. These representations are determined by both the objective features of the events or objects as well as the characteristics of the individual perceiving them including his or her goals or cultural values (Rosch, 1978; Shaver et al., 1987). Moreover, prototype theory proposes that these systems of mental representations consist of a vertical and a horizontal dimension (Rosch, 1978; Shaver et al., 1987). The former describes the hierarchical relation between those categories (e.g., vehicle - car – sports car). The horizontal dimension “concerns the segmentation of categories at the same level of inclusiveness" (Rosch, 1978, p. 30). Each of these segments, or categories, (“fuzzy sets”) is defined by a specific prototype, which is “an abstract image or set of features representing the best, most representative, most typical example of the category” (Shaver et al., 1987, p.

1062). The individual develops these prototypes through repeated exposure to the specific object or event of his or her environment. Accordingly, the individual categorizes new objects or events by comparing them with prototypes already acquired (Rosch, 1978; Shaver et al., 1987).

To study achievement emotions within the Rwandan context, it was intended to largely follow a process reported by Shaver et al. (1987) and Fontaine et al. (2002). First, a list of Kinyarwanda emotion terms was compiled. To do so, 150 students enrolled at a public university in Rwanda participated in a free-listing exercise. They all listed as many different emotion terms in Kinyarwanda as they could think of. In one of the next steps, a different sample of Rwandan psychology students from a local university rated these terms based on their similarity. Hierarchical cluster analysis was then used to turn the similarity data into an emotion taxonomy specific to the Rwandan context. It was hoped that an increasing disaggregation of the resulting cluster structure would result in identification of emotion prototypes such as enjoyment, anger, sadness. An emotion prototype is a cluster of different emotion terms that all relate to the same overarching emotional phenomenon (e.g. Fontaine et al., 2002; Shaver et al., 1987). For example, the cluster bringing together different emotion terms such as enthusiasm, amusement, contentment could be labelled enjoyment (Shaver et al., 1987). In a third step, it was then planned to let primary and secondary students rate the typicality of each emerging emotion prototype within an education context. In a final step, frequently occurring emotion prototypes would then be described by a different sample of school students, which in turn would then have provided the qualitative material to construct self-report scales to those achievement emotions most common within the Rwandan context (Shaver et al., 1987).

Figure 20: Hierarchical (average linkage) cluster analysis (pilot phase)



The free-listing exercise resulted in the identification of 117 Kinyarwanda emotions terms. The sorting of these terms, however, resulted only in weakly clustered data. Figure A2 on the next page presents the cluster diagram. At the top level, emotion terms can indeed be grouped into positively (e.g., to be encouraged) and negatively (e.g., to be angry) experienced emotions. It also appears that there is a division between emotions that are perceived to be negative by the person experiencing them (e.g., to be disgusted) as well as emotions that are perceived to be negative by those that are around the one experiencing them (e.g., pride/ vanity). Below this level, however, the interpretation of clusters became increasingly difficult. This may have been due to the fact that clusters were not well separated. Because of the poor cluster structure, it was then not possible to identify a manageable set of emotion prototypes that could have been used for any subsequent frequency analysis. Any grouping resulted either in a set of too few emotion prototypes or in a set of too many prototypes.

One reason for weakly clustered data could be low levels of agreements amongst respondents on how to best sort the emotion terms. A similar finding emerged in attempts to adapt the GRID questionnaire in Setswana, a language spoken by black native communities in the North-West Province of South Africa (Jonker et al., 2013). The GRID questionnaire is a tool for examining semantic concepts of various emotions across cultures (Fontaine et al., 2013). Despite extensive piloting efforts, the adaption to Setswana resulted in a tool that did not yield a clear representation of the meaning structure of emotion terms studied. Similar to the cluster analysis presented above, they only found a valence factor that distinguished between positive to negative emotion terms. According to the researcher, the conceptual structures were not clear "because Setswana participants agreed only moderately on the meaning of the emotion terms" (Jonker et al., 2013, p. 460). Within the current case, the poor interrater agreement could have been due to students' poor familiarity with the concept of emotions. In fact, during informal feedback interviews after the sorting exercise a number of study participants explained that they found the task interesting “because we never think about emotions”. This is remarkable given that all study participants were psychology students but may also highlight how uncommon it is to reflect about emotion terms and their meaning within Kinyarwanda amongst Rwandans. This all is not to say that the systematic study of emotions in Kinyarwanda is not possible. On the contrary, in line with Jonker et al. (2013) it is argued that additional efforts are necessary to explore how current methods to study important psychological constructs such as emotions are best studied in novel contexts such as Rwanda. However, it was deemed beyond the scope of this DPhil project to address these questions. It was for this reason that the framework of achievement emotions to study the drivers of student engagement in Rwanda was dropped. It was replaced by a broader focus on student motivation, a multidimensional construct that also allows to incorporate students’ emotional experiences. The way it was operationalized is explained in particular in section 3.

Annex 2: The (lack of) motivation and (lack of) confidence questionnaires (manuscript one)

The open-ended questionnaire on confidence

We need your help! All students have subjects they feel particularly confident about. For example, some students are particularly confident about themselves in Mathematics or Kinyarwanda. Maybe some students are particularly confident about themselves in social studies.

Can you help us to find out what confident students may think when they are studying in class? Imagine in class you are working on an activity given by the teacher, and you feel very confident about this exercise. What would you think? On the next page, please write down as many different thoughts you may have when you feel very confident about an exercise. On the next page, we gave you sentences to complete to make it easier for you. Write down whatever you believe confident students may think when working on exercises in class. Do not worry about right or wrong. This is no test.

Before you try to answer this question, please tell us more about yourself and the subject you are especially confident about.

	Name	Age	
Who are you?			Female Male
What subject are you especially confident about?			

When working on an exercise that I am confident about I think/ feel that…

When working on an exercise that I am confident about I think/ feel that…

When working on an exercise that I am confident about I think/ feel that…

When working on an exercise that I am confident about I think/ feel that…

When working on an exercise that I am confident about I think/ feel that…

When working on an exercise that I am confident about I think/ feel that…

The open-ended questionnaire on demotivation.

We need your help! All students have subjects they feel particularly demotivated about. For example, some students are particularly demotivated about Mathematics or Kinyarwanda. Maybe some students are particularly demotivated for social studies.

However, what are the reasons that we are particularly demotivated about some school subjects at school? Imagine in class you are working on an activity given by the teacher, and you are particularly demotivated about this exercise. On the next page, please write down as many different reasons why you may be demotivated about this exercise. On the next page, we gave you sentences to complete to make it easier for you. Write down whatever reasons for being demotivated to work on an exercise in class you can think of. Do not worry about right or wrong. This is no test.

Before you try to answer this question, please tell us more about yourself and the subject you are especially demotivated about.

	Name	Age	
Who are you?			Female Male
What subject are you especially demotivated about?			

I am demotivated to work on an exercise because…

I am demotivated to work on an exercise because…

I am demotivated to work on an exercise because…

I am demotivated to work on an exercise because…

I am demotivated to work on an exercise because…

I am demotivated to work on an exercise because…

The open-ended questionnaire on lack of confidence.

We need your help again! All students have subjects they feel particularly unconfident about. For example, some students are particularly unconfident about themselves in Mathematics or Kinyarwanda. Maybe some students are particularly unconfident about themselves in social studies.

Can you help us to find out what unconfident students may think when they are studying in class? Imagine in class you are working on an activity given by the teacher, and you feel very unconfident about this exercise. What would you think? On the next page, please write down as many different thoughts you may have when you feel very unconfident about an exercise. On the next page, we gave you sentences to complete to make it easier for you. Write down whatever you believe unconfident students may think when working on exercises in class. Do not worry about right or wrong. This is no test.

Before you try to answer this question, please tell us more about yourself and the subject you are especially unconfident about.

	Name	Age	
Who are you?			Female Male
What subject are you especially unconfident about?			

When working in class on an exercise I am unconfident about then I think that…

When working in class on an exercise I am unconfident about then I think that…

When working in class on an exercise I am unconfident about then I think that…

When working in class on an exercise I am unconfident about then I think that…

When working in class on an exercise I am unconfident about then I think that…

When working in class on an exercise I am unconfident about then I think that…

The open-ended questionnaire on motivation.

We need your help! All students have subjects they feel particularly motivated about. For example, some students are particularly motivated about Mathematics or Kinyarwanda. Maybe some students are particularly motivated for social studies.

However, what are the reasons that we are particularly motivated about some school subjects at school? Imagine in class you are working on an activity given by the teacher, and you are particularly motivated about this exercise. On the next page, please write down as many different reasons why you may be motivated about this exercise. On the next page, we gave you sentences to complete to make it easier for you. Write down whatever reasons for being motivated to work on an exercise in class you can think of. Do not worry about right or wrong. This is no test.

Before you try to answer this question, please tell us more about yourself and the subject you are especially motivated about.

	Name	Age	
Who are you?			Female Male
What subject are you especially motivated about?			

I am motivated to work on an exercise because…

I am motivated to work on an exercise because…

I am motivated to work on an exercise because…

I am motivated to work on an exercise because…

I am motivated to work on an exercise because…

I am motivated to work on an exercise because…

Annex 3: The subject-specific motivational self-report scales (manuscript two/ three)

Annex 3.1. The Scales on subject-specific motivational beliefs in chemistry and mathematics

Expectancy:

AB2	Do you understand [SUBJECT]?	Often	Sometimes	Never
AB2	Ese [SUBJECT] uraryumva	Akenshi	Rimwe na rime	Nta na rimwe

AB3	Do you do well in [SUBJECT]?	Often	Sometimes	Never
AB3	Ese mu [SUBJECT] ukora neza?	Akenshi	Rimwe na rime	Nta na rimwe

AB4	Do you know the [SUBJECT]?	Often	Sometimes	Never
AB4	Ese wumva uzi [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

AB5	Do you get good marks in [SUBJECT]?	Often	Sometimes	Never
AB5	Ese ugira amanota meza mu [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

AB6	Are you capable of [SUBJECT]?	Often	Sometimes	Never
AB6	Ese [SUBJECT] ura[ri]shobora?	Akenshi	Rimwe na rime	Nta na rimwe

AB8	Do you understand [SUBJECT]?	Often	Sometimes	Never
AB8	Ese [SUBJECT] ura[ri]sobanukirwa?	Akenshi	Rimwe na rime	Nta na rimwe

PO2	Do you like/ love this [SUBJECT]?	Often	Sometimes	Never
PO2	Ese ukunda [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

Costs:

AB7	Do you fail in [SUBJECT]?	Often	Sometimes	Never
AB7	Ese [SUBJECT] rirakunanira?	Akenshi	Rimwe na rime	Nta na rimwe

EF2	Is the subject of [subject] difficult?	Often	Sometimes	Never
EF2	Ese [SUBJECT] [ri]rakomera?	Akenshi	Rimwe na rime	Nta na rimwe

EF3	Is the subject of [subject] hard?	Often	Sometimes	Never
EF3	Ese [SUBJECT] [ri]ragora?	Akenshi	Rimwe na rime	Nta na rimwe

EF5	Is the subject of [subject] tiring?	Often	Sometimes	Never
EF5	Ese [subject] [ri]rarushya?	Akenshi	Rimwe na rime	Nta na rimwe

NE2	Does this [SUBJECT] make you sad?	Often	Sometimes	Never
NE2	Ese [SUBJECT] rirakubabaza?	Akenshi	Rimwe na rime	Nta na rimwe

NE4	Does this [SUBJECT] make you scared?	Often	Sometimes	Never
NE4	Ese [SUBJECT] rigutera kugira ubwoba?	Akenshi	Rimwe na rime	Nta na rimwe

Unimportance:

PI1	Do you think you don't give any value to [SUBJECT]?	Often	Sometimes	Never
PI1	Ese [SUBJECT] ntabwo u[yi]ha agaciro?	Akenshi	Rimwe na rime	Nta na rimwe

PI2	Do you think [SUBJECT] is useless for you?	Often	Sometimes	Never
PI2	Ese [SUBJECT] nta kamaro [ri]fite?	Akenshi	Rimwe na rime	Nta na rimwe

PI3	Do you think [SUBJECT] has no use for you?	Often	Sometimes	Never
PI3	Ese [SUBJECT] ntacyo [ri]kumariye?	Akenshi	Rimwe na rime	Nta na rimwe

PI4	Do you think [SUBJECT] is not getting you anywhere?	Often	Sometimes	Never
PI4	Ese [SUBJECT] ntacyo [ri]zakugezaho?	Akenshi	Rimwe na rime	Nta na rimwe

Value:

DA1	Does [SUBJECT] help you in your daily life?	Often	Sometimes	Never
DA1	Ese [SUBJECT] [ri]gufasha mu buzima bwa buri munsi?	Akenshi	Rimwe na rime	Nta na rimwe

DA2	Do you use [SUBJECT] in your everyday life?	Often	Sometimes	Never
DA2	Ese ukoresha [SUBJECT] mu buzima bwa buri munsi?	Akenshi	Rimwe na rime	Nta na rimwe

DA3	Is [SUBJECT] useful for your everyday life?	Often	Sometimes	Never
DA3	Ese [SUBJECT] [ri]kugirira akamaro mu buzima busanzwe?	Akenshi	Rimwe na rime	Nta na rimwe

GE1	Will [SUBJECT] help you in your future life?	Often	Sometimes	Never
GE1	Ese [SUBJECT] rizagufasha mu buzima bw'ejo.	Akenshi	Rimwe na rime	Nta na rimwe

GE2	Is [SUBJECT] useful for your future?	Often	Sometimes	Never
GE2	Ese [SUBJECT] rifitiye ahazaza hawe akamaro?	Akenshi	Rimwe na rime	Nta na rimwe

GE3	Is [SUBJECT] good for your future?	Often	Sometimes	Never
GE3	Ese [SUBJECT] ni ryiza ku hazaza hawe?	Akenshi	Rimwe na rime	Nta na rimwe

GE4	Will [SUBJECT] help you in what you will be?	Often	Sometimes	Never
GE4	Ese [SUBJECT] rizagufasha mu cyo uzaba cyo?	Akenshi	Rimwe na rime	Nta na rimwe

GE5	Will [SUBJECT] help you to get a good future?	Often	Sometimes	Never
GE5	Ese [SUBJECT] rizagufasha kuguhesha ejo heza?	Akenshi	Rimwe na rime	Nta na rimwe

Annex 3.2. The Scales on subject-specific motivational beliefs in English and Kinyarwanda

Expectancy:

AB2	Do you understand [SUBJECT]?	Often	Sometimes	Never
AB2	Ese [SUBJECT] uraryumva	Akenshi	Rimwe na rime	Nta na rimwe

AB3	Do you do well in [SUBJECT]?	Often	Sometimes	Never
AB3	Ese mu [SUBJECT] ukora neza?	Akenshi	Rimwe na rime	Nta na rimwe

AB4	Do you know the [SUBJECT]?	Often	Sometimes	Never
AB4	Ese wumva uzi [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

AB5	Do you get good marks in [SUBJECT]?	Often	Sometimes	Never
AB5	Ese ugira amanota meza mu [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

AB6	Are you capable of [SUBJECT]?	Often	Sometimes	Never
AB6	Ese [SUBJECT] ura[ri]shobora?	Akenshi	Rimwe na rime	Nta na rimwe

AB8	Do you understand [SUBJECT]?	Often	Sometimes	Never
AB8	Ese [SUBJECT] ura[ri]sobanukirwa?	Akenshi	Rimwe na rime	Nta na rimwe

Cost:

EF2	Is the subject of [subject] difficult?	Often	Sometimes	Never
EF2	Ese [SUBJECT] [ri]rakomera?	Akenshi	Rimwe na rime	Nta na rimwe

EF3	Is the subject of [subject] hard?	Often	Sometimes	Never
EF3	Ese [SUBJECT] [ri]ragora?	Akenshi	Rimwe na rime	Nta na rimwe

EF5	Is the subject of [subject] tiring?	Often	Sometimes	Never
EF5	Ese [subject] [ri]rarushya?	Akenshi	Rimwe na rime	Nta na rimwe

NE1	Do you hate this [SUBJECT]?	Often	Sometimes	Never
NE1	Ese wanga [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

NE2	Does this [SUBJECT] make you sad?	Often	Sometimes	Never
NE2	Ese [SUBJECT] rirakubabaza?	Akenshi	Rimwe na rime	Nta na rimwe

NE3	Does this [SUBJECT] discourage you?	Often	Sometimes	Never
NE3	Ese [SUBJECT] rigutera gucika intege?	Akenshi	Rimwe na rime	Nta na rimwe

NE4	Does this [SUBJECT] make you scared?	Often	Sometimes	Never
NE4	Ese [SUBJECT] rigutera kugira ubwoba?	Akenshi	Rimwe na rime	Nta na rimwe

OE2	Are you too busy to study [SUBJECT]?	Often	Sometimes	Never
OE2	Ese uba uhuze cyane kuburyo bikubuza kwiga [SUBJECT]?	Akenshi	Rimwe na rime	Nta na rimwe

Unimportance:

PI2	Do you think [SUBJECT] is useless for you?	Often	Sometimes	Never
PI2	Ese [SUBJECT] nta kamaro [ri]fite?	Akenshi	Rimwe na rime	Nta na rimwe

PI3	Do you think [SUBJECT] has no use for you?	Often	Sometimes	Never
PI3	Ese [SUBJECT] ntacyo [ri]kumariye?	Akenshi	Rimwe na rime	Nta na rimwe

PI4	Do you think [SUBJECT] is not getting you anywhere?	Often	Sometimes	Never
PI4	Ese [SUBJECT] ntacyo [ri]zakugezaho?	Akenshi	Rimwe na rime	Nta na rimwe

Value:

DA1	Does [SUBJECT] help you in your daily life?	Often	Sometimes	Never
DA1	Ese [SUBJECT] [ri]gufasha mu buzima bwa buri munsi?	Akenshi	Rimwe na rime	Nta na rimwe

DA3	Is [SUBJECT] useful for your everyday life?	Often	Sometimes	Never
DA3	Ese [SUBJECT] [ri]kugirira akamaro mu buzima busanzwe?	Akenshi	Rimwe na rime	Nta na rimwe

GE1	Will [SUBJECT] help you in your future life?	Often	Sometimes	Never
GE1	Ese [SUBJECT] rizagufasha mu buzima bw’ejo.	Akenshi	Rimwe na rime	Nta na rimwe

GE3	Is [SUBJECT] good for your future?	Often	Sometimes	Never
GE3	Ese [SUBJECT] ni ryiza ku hazaza hawe?	Akenshi	Rimwe na rime	Nta na rimwe

GE4	Will [SUBJECT] help you in what you will be?	Often	Sometimes	Never
GE4	Ese [SUBJECT] rizagufasha mu cyo uzaba cyo?	Akenshi	Rimwe na rime	Nta na rimwe

GE5	Will [SUBJECT] help you to get a good future?	Often	Sometimes	Never
GE5	Ese [SUBJECT] rizagufasha kuguhesha ejo heza?	Akenshi	Rimwe na rime	Nta na rimwe

Annex 4: The lesson-specific motivational self-report scales (manuscript three/ four/ five)

Daily Utility

DA2	Do you think you will use what you learned in today's activities in [subject] in your everyday life?	Yes, very much.	A bit	No, not at all.
DA2	Ese ibyo mwize mu bikorwa byakozwe muri iyi saha y' [SUBJECT] urumva uzabikorehsa mu buzima bwa buri munsi?	Yego cyane rwose.	Gake	Oya habe na gato.

DA1	Do you think today's activities of [SUBJECT] will help you in your daily life?	Yes, very much.	A bit	No, not at all.
DA1	Ese ibyo mwize mu bikorwa byakozwe muri iyi saha y' [SUBJECT] urumva bizagufasha mu buzima bwa buri munsi?	Yego cyane rwose.	Gake	Oya habe na gato.

DA3	Do you think today's activities of [subject] will be useful for your everyday life?	Yes, very much.	A bit	No, not at all.
DA3	Ese ibyo mwize mu bikorwa byakozwe muri iyi saha y' [SUBJECT] urumva bizakugirira akamaro mu buzima busanzwe?	Yego cyane rwose.	Gake	Oya habe na gato.

Effort

EF3	Were today's activities of [SUBJECT] hard?	Yes, very much.	A bit	No, not at all.
EF3	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] byari bigoye?	Yego cyane rwose.	Gake	Oya habe na gato.

EF5	Were today's activities of [SUBJECT] tiring?	Yes, very much.	A bit	No, not at all.
EF5	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] byari biruhije?	Yego cyane rwose.	Gake	Oya habe na gato.

EF6	Were today's activities of [SUBJECT] confusing?	Yes, very much.	A bit	No, not at all.

EF6	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] byari bicanganye?	Yego cyane rwose.	Gake	Oya habe na gato.

Expectancy

EX2	Do you feel you understood today's activities of [SUBJECT]?	Yes, very much.	A bit	No, not at all.
EX2	Ese ibyo mwize mu bikorwa byakozwe muri iyi saha y' [SUBJECT] urumva wabyumvise?	Yego cyane rwose.	Gake	Oya habe na gato.

EX5	Do you feel you were capable to do today's activities of [SUBJECT]?	Yes, very much.	A bit	No, not at all.
EX5	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] urumva wabishoboye?	Yego cyane rwose.	Gake	Oya habe na gato.

EX7	Do you feel you understood today's activities of [SUBJECT]?	Yes, very much.	A bit	No, not at all.
EX7	Ese ibyo mwize mu bikorwa byakozwe muri iyi saha y' [SUBJECT] urumva wabisobanukiwe?	Yego cyane rwose.	Gake	Oya habe na gato.

Unimportance

PI2	Do you think today's activities of [subject] were useless?	Yes, very much.	A bit	No, not at all.
PI2	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] wumvise nta kamaro?	Yego cyane rwose.	Gake	Oya habe na gato.

PI3	Do you think today's activities of [subject] had no use for you?	Yes, very much.	A bit	No, not at all.
PI3	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] wumvise ntacyo bimaze?	Yego cyane rwose.	Gake	Oya habe na gato.

PI1	Do you think you did not give any value to today's activities of [subject]?	Yes, very much.	A bit	No, not at all.
PI1	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] wumvise utabihaye agaciro?	Yego cyane rwose.	Gake	Oya habe na gato.

School Utility

SU2	Do you feel today's activities helped you prepare for the [SUBJECT] exams?	Yes, very much.	A bit	No, not at all.
SU2	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] urumva byagufashije kuzategura isuzumabumenyi?	Yego cyane rwose.	Gake	Oya habe na gato.

SU3	Do you feel today's activities helped you to do the exam of [SUBJECT]?	Yes, very much.	A bit	No, not at all.
SU3	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] urumva byagufashije kuzakora ikizamini cy'[SUBJECT]?	Yego cyane rwose.	Gake	Oya habe na gato.

SU1	Do you feel today's activities helped you succeed in [SUBJECT]	Yes, very much.	A bit	No, not at all.
SU1	Ese ibikorwa byakozwe muri iyi saha y' [SUBJECT] urumva byagufashije kuzatsinda [SUBJECT]?	Yego cyane rwose.	Gake	Oya habe na gato.

MISCELLANEOUS

M1	Did you learn anything new during the activities in [subject]	Yes, very much.	A bit	No, not at all.
M1	Ese hari igishya wize mubikorwa byakozwe muri iyi saha y'[subject]	Yego cyane rwose.	Gake	Oya habe na gato.

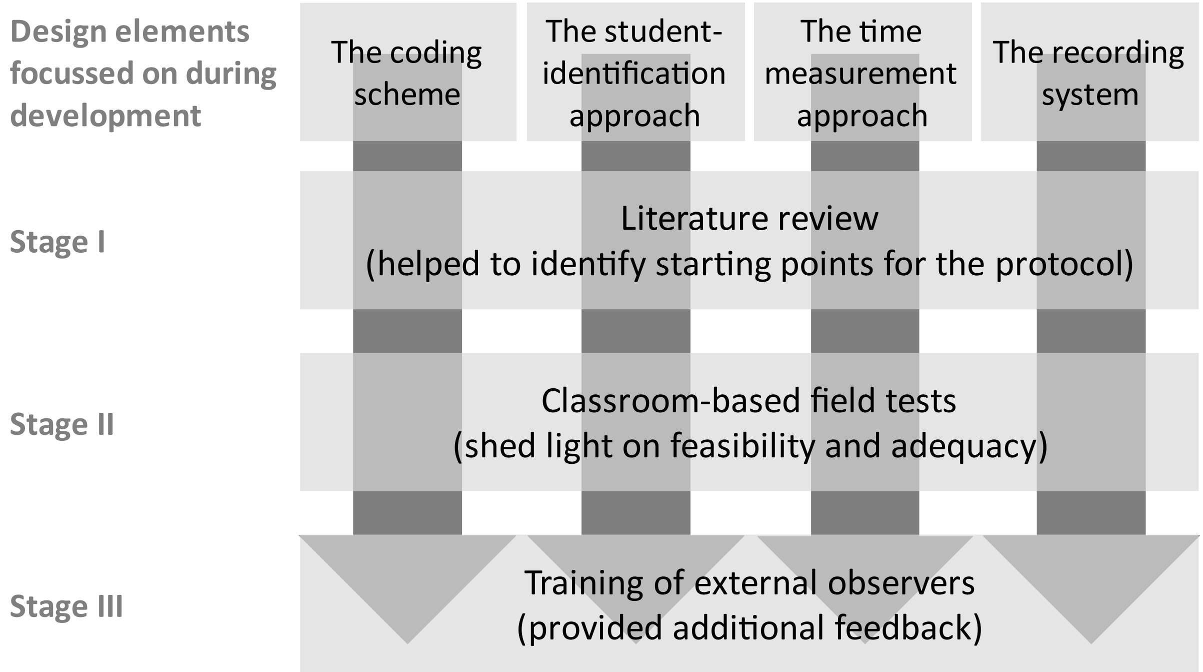
M2	Did you have any breakfast this morning?	Yes	No
M2	Ese wafashe ifunguro rya mugitondo?	Yego	Oya

Annex 5: The development of the observation protocol (manuscript five)

Whereas student motivation describes a person-internal phenomenon, student engagement refers to students' learning behaviours exhibited in the classroom. Unlike motivational experiences, these behaviours are to a large part directly "observable" (Skinner & Pitzer, 2012, p. 24). For example, whether or not a student is making a verbal contribution (i.e., agentic engagement) at any given point in time during a lesson can be inferred by observing this student during that particular moment. Since they are observable, the measurement of student behaviours does not necessarily require students to report on them. Hence, the development of locally grounded measurement instruments that ensure researchers and respondents interpret self-report items and their answers in the same manner is not necessary. It is for this reason that a development process for the student-engagement process was chosen that is best described deductively rather than inductively as in the case of student motivation. It is deductive in that the empirical operationalization of student engagement was derived from the wider literature and then adapted through classroom-based fieldwork especially during the months May through December 2019. This development process is outlined below. Figure A3 provides a schematic overview of the key aspects of the development of the student engagement process.

The tool-development process was governed by the needs the observation tool to be developed had to meet. Unlike other observation tools employed in Rwanda (Carter et al., 2020) or elsewhere (e.g., Abadzi, 2007; Adelman et al., 2015; Venäläinen, 2008), its main purpose was to facilitate repeated ratings of individual students during real-life lessons. Due to budget constraints, it was not possible to rely on video recordings of lessons that have partly been employed in other studies. Instead, external observers had to be located in class to be able to carry out student observations. Hence, it was necessary not only to develop a coding scheme that adequately covers the range of learning behaviours students at public lower-secondary schools in Rwanda commonly exhibit.

Figure 21: Schematized overview of the development process of the observation tool.



It was also necessary to develop a practical student-identification system that can be employed during real-life lessons to spot those pre-determined students that have to be observed and rated subsequently. In the context of Rwanda, this challenge was further exacerbated by the fact that students are generally in the habit of changing the location in class where they are seated between lessons. Also, since student ratings had to be carried out repeatedly by different external observers across different classrooms, a time-tracking devise had to be identified and incorporated into the observation protocol in a way that it ensures all observations are comparable in terms of their length and timing. Finally, a recording system had to be developed that allowed external observers to record their observation results while being located in class. In order to find solutions for each of the four points mentioned, the literature was reviewed first. Then potential solutions identified were field tested in real-life classrooms. All solutions were then spelt out in the written protocol summarizing the student procedures. During the theoretical and practical training sessions, this protocol was then subject to extensive discussions with observers that were hired to carry out the classroom observations later on during the main project phase. These discussions helped to refine various aspects of the protocol.

In order to develop a coding scheme to categorize students' behavioural engagement, the following steps were taken. First, the wider educational literature was reviewed in order to draft a coding scheme on student engagement for the Rwanda context. In this respect, the protocols reported by Ocumpaugh et al. (2015) and Shapiro (2011) were especially of interest. Both provide detailed accounts of how behavioural engagement at the individual level is operationalized and how the observation procedures are carried out. The tool described by Ocumpaugh et al. (Ocumpaugh et al., 2015) focusses on both behavioural and emotional engagement. The latter was not of interest to the current research. The former is operationalized amongst others through the categories of on-task behaviour, which refers to "a student who is doing what he or she is supposed to be doing" as well as on-task conversation, which refers to "student who is working towards his or her assignment while having a conversation with the teacher or another student about the subject matter or learning task" (p. 38). Shapiro's (2011) tool also distinguishes between two engagement states. Active engagement implies amongst others students talking to the teacher or a peer about the assigned material whereas passive engagement implies students being passively attending to classroom work. Also, both coding schemes provide different options for categorizing disengagement.

Once the coding scheme was drafted, it was field tested within lower-secondary classes at schools outside Kigali. For that purpose, two schools were recruited. All targeted classes were visited, and the research purpose was introduced. All students were given consent forms for them to review as well as for their parents. All those that returned duly signed consent forms were included into the research. At this point within the research project, several pilot work streams were carried and thus testing the draft coding scheme was only one. To test it, different classes as well as different lessons within classes were attended. At this point, this DPhil student closely worked with two female Rwanda research assistants both fluent in Kinyarwanda and English. On each field day any of the two accompanied this DPhil student during school visits. Both were trained up on the coding scheme before field work started. During lessons attended, both coders randomly selected a group out of all participating students, which in turn were repeatedly observed and rated during class using the draft coding scheme. Ratings were written down along with some comments on a given a rating was selected. Upon completion of classroom observations, both raters then compared results and discussed interpretations.

Joint discussions of coding results helped to refine the coding scheme both in terms of the conceptualizations of the different categories as well as their empirical operationalization. These joint discussions also helped to finalize other important aspects of the observation protocol such as the length of student observations. Initially, observations of individual students were supposed to last for one second only (i.e., momentary time-sampling approaches). Continuous piloting and discussions with research assistants showed that a slightly longer interval resulted in better coding outcomes. It was for this reason that student observations were pegged at five seconds. Another important finding that emerged from those school-based trials was the need to control for classroom activities when rating student engagement. Classroom observations revealed that students behave differently when an actual task is given compared to when teachers are unresponsive to students learning. These insights eventually resulted in the decision to rate classroom activities during student observations as well. To code classroom activities, the World Bank's (2015) revised version of Stallings' (e.g., 1977) Classroom Snapshot Observation System was adjusted to the Rwandan context, a system that has already been applied to the developing-country contexts (e.g., Abadzi, 2007; Adelman et al., 2015; Venäläinen, 2008). Similar to the development of the engagement coding scheme, the classroom-activity coding school was field tested during real-life classes until a version emerged that appeared to adequately cover those classroom activities commonly found in Rwandan classrooms. The final coding schemes employed are presented in table 4 and 5.

Classroom observations of all students within a class was not possible. It was therefore obvious that a system needs to be put in place that allows an effective identification of students in class. A key challenge in this respect has been that in Rwanda lower-secondary classes were often large easily comprising of 50 or more students. Students within a class often change the location where they are seated. Moreover, since students are expected to clean their classrooms at the end of a day before going home, classroom benches are moved around almost daily. Hence, the bench arrangement within a classroom frequently changes. Several options to identify students were tested. At first, students were handed out tags that they could attach to the collar of their school uniforms. These tags were made out of paper and attached to the uniforms using safety pins. This, however, did not turn out to be a viable option. After one or two days, the number of students that lost their tags rapidly increased. Replacing these tags turned out to be logistically challenging especially when it had to be done on a daily basis. Also, a concern was that students may hurt themselves with the safety pins handed out to them. A second option was to take photos of students. To test this idea, research assistants were given a sample of photos and they then had to identify the students in class. Even when photos were printed with high resolution, this exercise turned out to be very difficult often resulting in false positives. Reading out student names in the way outlined in the methodology chapter turned out to be the most effective of student identification. It was therefore included into the protocol.

Protocols such as the ones reported by Ocumpaugh et al. (2015) and Shapiro (2011) generally rely on tablets to support external raters in their student observations. To do so, they require specific software to run on these tablets. Despite several attempts, it was not possible to identify any provider in Rwanda that could have provided a bespoke software package attuned to the needs of the emerging observation protocol at affordable cost. It was therefore decided to make the recording of student ratings paper based. A challenge in this respect was then how to ensure that all external raters follow the same time scheme. Otherwise, classroom-based observations by the support staff may have varied in terms of interval length or start and end time. To ensure all external observers follow the same time sampling, it was decided to equip all external observers with no-frills smart phones. A review of different options led to the identification of the multi-timer app provided by catfantom.[31] It is freeware, it allows running different timers in any order required, and gives audio signals when an interval is over. In order not to disturb classes through repeated peeping, all observers were equipped with earphones. The application of the multi-timer app during classroom-based observations was then field tested in the weeks around August through September 2019. At this point, different interval lengths for both observing classroom activities as well as observing student engagement were again tired out. An interval length of five seconds for both again turned out to be the most preferred option. Another feature that was included into the protocol at that time was the initial warming-up phase of three minutes right at the beginning of each lesson. Classroom observations revealed that students needed some minutes to settle in, getting out their learning materials before focussing on the lesson. The final time-sampling protocol adopted is outlined in the methodology chapter.

[31] https://play.google.com/store/apps/details?id=org.catfantom.multitimer&hl=en_US&gl=US.

Another issue that emerged due to the fact that no observation software was employed was the recording of the ratings during classroom observations. It had to be paper based. A key challenge in this respect was to design a form that allowed observers to record ratings while being able to operate the time-app on their smart phones in a way that is not obtrusive to teaching and learning in class. To achieve this, it was decided to equip observers with clipboards. In order to ensure observers can support recording form and smart phone at the same time, the surface of the clipboard was further extended through a solid carton paper sheet that was fastened to the board through its clip at the top. Different arrangements as well as different designs of the recording form were field tested. The final version is provided in annex 4. Throughout all those pilot stages that helped to finalize the student-identification procedure, the time-sampling design, and the recording of rating results, the continuous field testing of the two coding schemes remained an important aspect of the classroom work.

The final stage of the development of the engagement-observation tool concerned the training of external observers. In total, eight external observers (six females and two males) were hired and trained up. Due to health reasons, one of the observers quit the work upon completion of the training. To ensure an effective training, a detailed observation protocol was written up that spelt out all coding schemes as well as the student-identification and observation procedures. Its English version consists of 28 pages. This protocol was then translated into Kinyarwanda. To ensure quality of the translation, a first research assistant translated this tool into Kinyarwanda. A second research assistant then translated it back into English. The second English version was then compared with the English original. That way, it was able to check the Kinyarwanda translations of key aspects such as the coding schemes. The training of external observers in classroom observations consisted of several rounds. The first round included the theoretical training whereas subsequent training rounds became increasingly more practice based. The theoretical training lasted for five days and was carried out in October 2019. It covered several aspects. First, the coding schemes were discussed. Different exercises were used to help observers become increasingly more familiar with the coding categories. Next, the time management associated with classroom-based observations were introduced. Key in this respect was the operation of the multi-timer app. Afterwards, the student-identification procedures were explained. Key in this respect was the identification of students through calling out their names, marking their locations into the seating plan, and re-identification of each target student by the means of the seating plans repeatedly over some time. To practice these aspects, different classroom exercises were designed involving real people recruited for that purpose. Next, how to code observation results were practiced by the team. Throughout the entire theoretical training, all observers were repeatedly tested on the definitions and operationalizations of the different coding categories. That way, it was ensured that they all become increasingly familiar with the coding schemes and their applications.

The practical training consisted of three rounds carried out between October 2019 and January 2020. In total, it comprised of 19 days. Each round was carried out at different schools inside and outside Kigali. The first round consisted of five days and took place at a public school in one of districts outside Kigali in the second week of October 2019. The second round consisted of eight days in the second and third week of December 2019 and took place in one of the private schools in Kigali. Between October and January all public schools are closed in Rwanda. Since private schools in Rwanda often follow the international (i.e., Cambridge) curriculum, they also follow a different school calendar than public Rwandan schools. Since private schools in Rwanda constitute a rather different learning environment than public schools (e.g., smaller class sizes, teaching in English etc.), it was decided to carry out a final round of practical training early January 2020. It happened around the time when student recruitment for the main study of this DPhil project took place.

The different practical training rounds had slightly different objectives. The first round in October as well as the second round in December mainly served the purpose to provide external observers with practical experience in carrying out classroom observations following the procedures spelt out in the observation protocol. Through both rounds, this DPhil student accompanied each of the observers into class. Both carried out observations and ratings of students and classroom activities independently of each other. To be synchronized, they both followed the same time app by sharing headphones. After completion of a lesson, both shared coding results, determined percentage-based agreements as a sign of the aligned understanding coding schemes and observation procedures, and discussed potential areas of misunderstandings. In addition to these supervised classroom-observations, observers also carried out classroom-observations in team of two. The latter allowed them to gain additional practical experience in the observation procedures, to identify potential areas of confusions, and to provide each other with additional feedback. Especially through the percentage-based agreements it was possible to track those observers that required additional retraining. The third round of practical training mainly consisted of observers carrying out joint classroom observations. This in turn helped observers to refamiliarize themselves with all aspects of the observation procedure ahead of the main study that was carried out soon afterwards. The final round was completed by an end-of-training evaluation of each observer. For that purpose, this DPhil student attended classroom-based observations of each team member. Both jointly but independently rated student behaviours as well as classroom activities. Agreements were then determined using Cohen’s Kappa (Cohen, 1960).

Annex 6: Author contributions to the manuscripts (CRediT)

Using the contributor role taxonomy (Credit), the role of each author of each of the manuscripts are specified below.

Bulla, D., Uwitonze, A. G., Malmberg, L.-E.,, and Strand, S., (2022a). The cognitive-affective structure of "icyizere" (confidence) and "ishyaka" (motivation) among Rwandan students. A mixed-methods study. *International Journal of Educational Research*.

Author 1: Project administration, resources, conceptualization, methodology, investigation, data curation, formal analysis, software, writing—original draft. Author 2: investigation, data curation, formal analysis. Author 3: Supervision, visualization, writing—review and editing. Author 4: Supervision, writing—review and editing.

Bulla, D., Malmberg, L.-E., & Strand, S. (2022b). The Generalized Internal/External Frame of Reference Model among Secondary School Students in Rwanda. *Contemporary Educational Psychology*.

Author 1: Project administration, resources, conceptualization, methodology, investigation, data curation, formal analysis, software, writing—original draft. Author 2: Supervision, visualization, writing—review and editing. Author 3: Supervision, writing—review and editing.

Bulla, D., Malmberg, L.-E., & Strand, S. (2022c). The generalized internal/external frame of reference model with average motivational states. Findings from Rwanda. *Educational Research and Evaluation*.

Author 1: Project administration, resources, conceptualization, methodology, investigation, data curation, formal analysis, software, writing—original draft. Author 2: Supervision, visualization, writing—review and editing. Author 3: Supervision, writing—review and editing.

Bulla, D., Malmberg, L.-E., & Strand, S. (2022d). Does the time allocation across classroom activities predict motivational states? Findings from Rwandan secondary schools. *Learning and Individual Differences*.

Author 1: Project administration, resources, conceptualization, methodology, investigation, data curation, formal analysis, software, writing—original draft. Author 2: Supervision, visualization, writing—review and editing. Author 3: Supervision, writing—review and editing.

Bulla, D., Malmberg, L.-E., & Strand, S. (2022e). Is being on-task predicted by motivational self-reports and classroom activities? Results from an observational study in Rwanda. *Learing and instruction*.

Author 1: Project administration, resources, conceptualization, methodology, investigation, data curation, formal analysis, software, writing—original draft. Author 2: Supervision, visualization, writing—review and editing. Author 3: Supervision, writing—review and editing.

Annex 7: Supplemental material to manuscript one

Interrater reliabilities, agreements and deferrals rates.

` As explained in the article, coding was carried out in rounds. In case of both data sets, the first round was on the perceived relevance of a given data segment. A segment was considered irrelevant if it did not provide any information on experiences of (de-)motivation or (lack of) confidence (e.g., "I am motivated to work on an exercise because…I have motivation in my studies"). During the second round, the type of information provided by the segment was coded. A segment might have fallen into any of the 4 types: confidence, demotivation, lack of confidence, and motivation. The final coding in this respect did not always match the type of questionnaire the data segment was from. With round 3, main coding started. As explained in the main text, we first applied the main categories to the data. In case of (de-)motivation, main categories excluded the different facets of utility value as well as source and consequence. In the case of (lack of) confidence, this excluded type of negative affective reactions as well as source and consequence. Hence, during the first follow-up coding of the motivational data, we applied those utility-related sub-categories to those segments that were initially coded as pertaining to utility perceptions. During the second follow-up coding, we then coded those segments as source or consequence that were initially coded as miscellaneous. In case of the (lack of) confidence, there was only one follow-up coding round. Here, we applied the sub-categories negative and positive affective reactions to those segments in the main category pertaining to emotions and source or consequence to those segments in the miscellaneous category. Interrater reliabilities, final agreements, and deferral rates are presented on the next page.

Table 61: Interrater reliabilities, agreements and deferrals rates.

Round			Results: Motivation Data	Results: Confidence Data
1	Guiding question		Is the statement relevant?	Is the statement relevant?
		n	1071	1172
	Krippendorff's Alpha		0.39	0.46
	Interrater Agreement		100%	100%
	Deferrer	Coder 1	45.6%	52.4%
		Coder 2	54.4%	47.6%
		Neither	0%	0%
2	Guiding question		Is the statement about motivation or demotivation?	Is the statement about confidence or unconfidence?
		n	1017	1172
	Krippendorff's Alpha		0.93	0.83
	Interrater Agreement		100%	100%
	Deferrer	Coder 1	55%	37.6%
		Coder 2	45%	50.6%
		Neither	0%	11.8%
3	Guiding question		What cognitive-affective factor does the statement refer to?	What cognitive-affective factor does the statement refer to?
		n	995	1073
	Krippendorff's Alpha		0.68	0.51
	Interrater Agreement		100%	100%
	Deferrer	Coder 1	26.4%	12.2%
		Coder 2	49.1%	84.1%
		Neither	24.5%	3.7%
4	Guiding question		What utility facets do utility statements refer to?	Do statements refer to consequence, source, positive or negative affect?
		n	193	
	Krippendorff's Alpha		0.91	0.63
	Interrater Agreement		100%	100%
	Deferrer	Coder 1	13%	13.8%
		Coder 2	43.5%	71.6%
		Neither	43.5%	14.7%
5	Guiding question		Do statements refer to consequence or source?	
		n	225	
	Krippendorff's Alpha		.58	
	Interrater Agreement		100%	
	Deferrer	Coder 1	24.4%	
		Coder 2	66.7%	
		Neither	8.9%	

The qualitative content analysis of the data segments on "source" and "consequence".

Two categories that emerged during coding across both the (de)motivation and (un)confidence datasets were sources and consequences. We defined source as determinants of (de)motivation and (un)confidence other than cognitive-emotional factors discussed within the expectancy, value, and cost section of the main article. 42.8% of (de)motivation respondents and 33.3% of (un)confidence respondents provided respective information. They accounted for around 23.9% of the (de)motivation coding and just 7.6% of (un)confidence coding. We further sub divided source into altogether 6 sub-categories. Within (un)confidence, students explained the presence of confidence through study habits (53.1%). Study habits in this respect referred to ways to ensure that one has understood school content. To do so, students revert friends and teachers. For example, a male S3 student explained that "I might approach the teacher and get explanations of the things I did not understand well". Others pointed out that studying is an important driver of confidence. A male S1 student stated that "exercise is the key to confidence" because it ensures "that I might understand what I didn't know". Within motivation students tended to explain the presence of (de)motivation through study habits to a lesser extent (21.7%). For example, a female S1 student mentioned that she felt demotivated when "I might not have read for the given subject". In the case of (de)motivation, it was often difficult to determine of student answers refer to the source or consequence of (de)motivation. For example, another female S1 student mentioned she may feel demotivated "because I was not paying attention the time of the explanation". It might have implied that she didn't feel confident about certain topics because she didn't follow the teacher explanations previously. However, her statement might have also implied that she didn't pay attention to the teacher explanations because she felt demotivated even back then. Another source of confidence (19.6%) and motivation (19.4%) respectively mentioned by students was teaching quality. Teachers' abilities to explain content was deemed a determinant of subjective confidence as well as motivation. A female S3 student pointed out that "I have understood what the teacher taught which makes me have confidence in myself" whereas a female S1 student reported "the teacher scolds you, so you cannot be confident". A similar picture emerges within motivation. A female S1 student wrote “there are times the teacher doesn't explain it well, so your motivation decreases” whereas another S1 student stated “when the teacher gives me an exercise and I fail he explains it to me which [is] why I get motivated to do the exercise”. Another source of (un)confidence (18.6%) and (de)motivation (20%) identified was coined general school support. This sub-category was about general factors that influence students’ ability to study. One important aspect in this respect was that students wanted to have someone who could help them in their studies. For example, a female S1 student explained that “I feel if I had someone to explain well I would succeed”. An important resource here appeared to be friends as exemplified by the following statement of a male S2 student: “it makes me very happy to see that I can work with my friends”. This aspect also shined through within the sub-category of study habits. Study participants generally seemed to enjoy co-studying with their friends and classmates. General school support also referred to the kind of support parents render to their children. A female S2 student mentioned to be motivated because her parents “give me enough/sufficient school materials and pay for my school fees on time” while a female S1 student wrote “my parents encourage me to study”. Two other sub-categories that came up within motivation but hardly in confidence were the sub-categories of personal problems and situational factors (both 19.4% each). The former refers to challenges such as poverty or mental health. According to a male S2 student, for example, one may experience demotivation “because sometimes I don’t feel well, have a lot of problems in me”. Other students mentioned they feel demotivated because of “missing school materials” (male S2 student) such as “pens” and “books” (male S1 student). Situational factors are rather situation-specific aspects that may shape one’s experience of motivation. These factors include “being ill” (female S1 student), “feeling hungry” (male S3 student), or “feeling tired” (female S2 student).

As for consequence, fewer sub-categories were identified. Across confidence and motivation, they comprised expectations, engagement, and reflections. Engagement described especially the behavioural consequences of motivation and confidence. Within the motivation subset all data segments pertaining to consequences were coded that way. The behavioural consequences of demotivation were often described as “laziness” (female S1 student). A male S1 student mentioned that “I don't want to hear about what the teachers are telling me” when feeling demotivated. By contrast, motivation is associated with positive behavioural consequences. A male S1 student explained that motivation “helps me to work harder” whereas a female S3 student experienced that “it helps us to think more”, which may point at cognitive effects of motivation. Within the confidence subsets, around 30.8% of consequence-related data segments were coded as behavioural engagement. The types of behavioural consequences reported resembled those that were observed for (de)motivation. A male S1 student explained that because of confidence “I will work so hard and use all my intelligence so I can get good marks” whereas a female S1 student mentioned that due to confidence “I have the energy to do the exercise”. By contrast, unconfidence stifles behavioural engagement. For example, a female S1 student said that because of lack of confidence “I won’t do it [most likely, she was referring to a class exercise], I get discouraged and fell like not doing it”. The findings that confidence and motivation were associated with behavioural engagement whereas demotivation and unconfidence were associated with behavioural disengagement were taken as further evidence that both the terms *ishyaka* (motivation) and *icyizere* (confidence) were interpreted by Rwandan students in a way that resembled ours. The last two sub-categories that were identified within the consequence subset of the (un)confidence data were expectations (61.4%) and reflections (7.7%). Statements within the former sub-category often implied the notion that confidence helps with one’s studies, which in turn helps to acquire a bright future of support by others.

For example, a female S1 student explained that because of confidence “I would excel in the exercise I' m doing and that I will have a good future”. The final sub-category was labelled reflections. Statements here suggested that some students were contemplating especially about the hindrances towards academic success. For example, a female S1 student stated that “I wish to succeed but there are things I don’t understand”. Altogether only 5 answer statements were included into this sub-category.

Annex 8: Supplemental material to manuscript two

Table 62: Descriptive statistics for motivational self-report items (manuscript two)

	Chemistry						Math						English						Kinyarwanda					
	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs
												Expectancy Beliefs												
AB2	2.09	0.61	-0.04	2.67	0.10	0.02	2.19	0.63	-0.17	2.4	0.08	0.03	2.46	0.58	-0.49	2.29	0.14	0.01	2.60	0.55	-0.96	2.88	0.05	0.02
AB3	2.14	0.58	-0.01	2.84	0.13	0.03	2.23	0.6	-0.14	2.5	0.07	0.04	2.52	0.54	-0.46	1.96	0.09	0.02	2.58	0.55	-0.86	2.68	0.03	0.03
AB4	2.07	0.68	-0.09	2.18	0.08	0.02	2.17	0.67	-0.21	2.21	0.06	0.03	2.39	0.62	-0.49	2.35	0.11	0.03	2.49	0.61	-0.78	2.62	0.04	0.03
AB5	2.18	0.56	0.02	2.82	0.14	0.01	2.14	0.62	-0.09	2.55	0.08	0.03	2.49	0.53	-0.28	1.80	0.09	0.01	2.60	0.57	-1.04	3.08	0.09	0.01
AB6	2.11	0.6	-0.05	2.69	0.13	0.02	2.17	0.6	-0.10	2.58	0.09	0.02	2.46	0.57	-0.43	2.20	0.13	0.02	2.56	0.56	-0.82	2.64	0.03	0.02
AB8	2.13	0.61	-0.08	2.59	0.11	0.02	2.24	0.61	-0.19	2.43	0.08	0.03	2.49	0.59	-0.64	2.44	0.13	0.01	2.64	0.53	-1.08	3.09	0.03	0.02
PO2	2.44	0.69	-0.82	2.45	0.10	0.02	2.49	0.66	-0.94	2.74	0.05	0.02												
												Utility-Value Beliefs												
DA1	2.49	0.64	-0.86	2.68	0.02	0.01	2.71	0.53	-1.68	4.91	0.01	0.01	2.69	0.49	-1.12	2.97	0.02	0.03	2.82	0.42	-2.36	7.97	0.01	0.01
DA2	2.27	0.69	-0.40	2.14	0.02	0.01	2.63	0.57	-1.30	3.69	0.00	0.02												
DA3	2.46	0.64	-0.80	2.58	0.04	0.02	2.73	0.52	-1.77	5.24	0.01	0.02	2.71	0.51	-1.48	4.26	0.03	0.01	2.76	0.47	-1.69	4.93	0.00	0.01
GE1	2.62	0.61	-1.38	3.79	0.03	0.02	2.79	0.46	-2.18	7.04	0.01	0.01	2.84	0.39	-2.24	7.17	0.03	0.01	2.73	0.51	-1.75	5.21	0.01	0.02
GE2	2.53	0.65	-1.07	2.99	0.02	0.01	2.75	0.51	-1.91	5.82	0.00	0.02												
GE3	2.63	0.60	-1.39	3.85	0.01	0.03	2.79	0.47	-2.24	7.30	0.03	0.01	2.86	0.38	-2.77	10.35	0.01	0.01	2.69	0.53	-1.49	4.29	0.00	0.02
GE4	2.48	0.70	-1.00	2.66	0.02	0.01	2.7	0.54	-1.67	4.86	0.05	0.01	2.81	0.45	-2.33	7.81	0.00	0.02	2.53	0.66	-1.07	2.94	0.02	0.01
GE5	2.58	0.64	-1.26	3.41	0.01	0.02	2.77	0.48	-1.99	6.18	0.02	0.02	2.86	0.38	-2.58	9.12	0.00	0.01	2.64	0.58	-1.36	3.83	0.00	0.02

	Chemistry						Math						English						Kinyarwanda					
	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs	Mean	SD	SK	KU	ICC	MVs
Cost Beliefs																								
AB7	2.00	0.63	0.00	2.54	0.05	0.01	2.03	0.63	-0.02	2.54	0.06	0.01												
EF2	1.86	0.63	0.12	2.47	0.02	0.02	1.86	0.64	0.13	2.36	0.04	0.03	2.27	0.64	-0.31	2.30	0.05	0.02	2.24	0.70	-0.36	2.07	0.04	0.02
EF3	1.85	0.63	0.12	2.47	0.05	0.05	1.89	0.65	0.11	2.34	0.02	0.05	2.26	0.64	-0.3	2.29	0.06	0.02	2.25	0.70	-0.39	2.08	0.05	0.02
EF5	1.85	0.64	0.15	2.37	0.04	0.02	2.15	0.64	-0.14	2.37	0.05	0.01	2.28	0.65	-0.34	2.27	0.05	0.01	2.25	0.68	-0.35	2.17	0.03	0.02
NE1													2.63	0.66	-1.53	3.92	0.04	0.01	2.60	0.68	-1.43	3.60	0.01	0.03
NE2	2.29	0.74	-0.52	1.98	0.01	0.03	2.48	0.65	-0.88	2.68	0.04	0.01	2.51	0.70	-1.07	2.81	0.01	0.02	2.52	0.70	-1.13	2.90	0.02	0.01
NE3													2.54	0.66	-1.13	3.05	0.03	0.02	2.52	0.71	-1.12	2.84	0.02	0.02
NE4	2.14	0.75	-0.23	1.8	0.01	0.02	2.11	0.76	-0.19	1.73	0.02	0.02	2.52	0.67	-1.06	2.9	0.08	0.01	2.54	0.67	-1.14	3.05	0.02	0.01
OE2													2.34	0.74	-0.65	2.07	0.01	0.01	2.36	0.75	-0.69	2.07	0.03	0.02

Note: SD = standard deviation; SK = skewness; KU = kurtosis; ICC = Intra-cluster correlations; MVs = percentage of missing value

Table 63: Path coefficients from latent regressions of motivational beliefs on prior achievement (basic-education schools only; model 3; manuscript two)

Subset	Mean	SD	Skewness	kurtosis	ICC
Chemistry	48.78	14.24	0.37	2.94	0.53
Chemistry (day)	43.95	10.67	0.23	3	0.32
Chemistry (boarding)	54.23	15.75	-0.07	2.65	0.59
Chemistry (female)	47.47	12.85	0.16	2.98	0.54
Chemistry (male)	50.11	15.46	0.42	2.67	0.55
English	57.11	13.53	-0.42	3.23	0.64
English (day)	53.91	14.42	-0.21	3.24	0.72
English (boarding)	62.6	10.34	-0.61	4.08	0.51
English (female)	57.21	12.66	-0.42	3.18	0.69
English (male)	58.78	14.08	-0.67	3.79	0.66
Kinyarwanda	56.77	16.15	-0.63	2.6	0.65
Kinyarwanda (day)	52.37	18.9	-0.2	1.8	0.71
Kinyarwanda (boarding)	62.15	11.16	-0.82	4.24	0.47
Kinyarwanda (female)	57.17	15.89	-0.67	2.53	0.62
Kinyarwanda (male)	56.71	17.11	-0.64	2.55	0.68
Mathematics	43.98	15.74	0.37	2.72	0.48
Mathematics (day)	37.06	13.13	0.78	3.98	0.18
Mathematics (boarding)	52.55	14.62	-0.03	2.92	0.64
Mathematics (female)	43.31	13.72	0.21	2.64	0.45
Mathematics (male)	45.36	17.77	0.35	2.42	0.52

Note. ‘day’ refers to average school marks from day schools. ‘boarding’ refers to average school marks from boarding schools. ‘female’ refers to average school marks from females. ‘male’ refers to average school marks from males.

Table 64: Path coefficients from latent regressions of motivational beliefs on prior achievement (basic-education schools only; model 3; manuscript two)

Predictor	Expectancies	Value	Costs	Predictions
		Chemistry outcomes		
Chemistry	0.191*	0.079	0.006	Match (++)
English	-0.069	0.003	0.026	Far (-)
Kinyarwanda	0.054	0.100	-0.003	Far (-)
Math	0.118	0.093	0.152**	Near (+)
		English outcomes		
Chemistry	-0.012	0.078	-0.140*	Far (-)
English	0.027	0.044	0.216***	Match (++)
Kinyarwanda	0.070	0.108*	0.133***	Far (-)
Math	-0.013	-0.038	0.006	Far (-)
		Kinyarwanda outcomes		
Chemistry	-0.151*	0.036	-0.211***	Far (-)
English	0.049	0.008	0.152**	Far (-)
Kinyarwanda	0.250***	0.103	0.139***	Match (++)
Math	-0.002	0.028	0.065	Far (-)
		Math outcomes		
Chemistry	0.037	0.088	-0.040	Near (+)
English	-0.009	-0.042	0.057	Far (-)
Kinyarwanda	-0.063	0.020	-0.101**	Far (-)
Math	0.502***	0.173**	0.341***	Match (++)

Note. All parameters are standardized. ***p < .001, **p < .01, *p < .05.

Table 65: Path coefficients from latent regressions of motivational beliefs on prior achievement (boarding schools only; model 4; manuscript two)

Predictor	Expectancies	Value	Costs	Predictions
		Chemistry outcomes		
Chemistry	0.665***	0.313**	0.628***	Match (++)
English	-0.109	-0.071	-0.159	Far (-)
Kinyarwanda	-0.224***	0.023	-0.131**	Far (-)
Math	-0.063	-0.072	-0.074	Near (+)
		English outcomes		
Chemistry	-0.237***	-0.067	-0.004	Far (-)
English	0.586***	0.199***	0.360***	Match (++)
Kinyarwanda	-0.029	0.100	-0.106	Far (-)
Math	-0.388***	-0.064	-0.199**	Far (-)
		Kinyarwanda outcomes		
Chemistry	-0.218***	-0.046	-0.165	Far (-)
English	-0.137*	-0.049	0.000	Far (-)
Kinyarwanda	0.536***	0.250***	0.267***	Match (++)
Math	-0.157*	-0.057	-0.049	Far (-)
		Math outcomes		
Chemistry	0.085	0.112	0.157*	Near (+)
English	-0.252***	0.058	-0.110	Far (-)
Kinyarwanda	-0.156**	0.020	-0.132*	Far (-)
Math	0.592***	0.084	0.397***	Match (++)

Note. All parameters are standardized. ***p < .001, **p < .01, *p < .05.

Table 66: Path coefficients from the multiple-group regression models grouped by school type and controlling for gender (manuscript two)

Predictor	Expectancies		Value		Costs		Predictions
	Day school	Boarding school	Day school	Boarding school	Day school	Boarding school	
			Chemistry outcomes				
Chemistry	0.191*	0.650***	0.069	0.320**	0.022	0.609***	Match (++)
English	-0.069	-0.111	0.002	-0.070	0.024	-0.160	Far (-)
Kinyarwanda	0.056	-0.223***	0.096	0.022	-0.001	-0.132**	Far (-)
Math	0.120	-0.066	0.087	-0.070	0.156*	-0.078	Near (+)
Gender	-0.056	-0.073	0.110	0.022	-0.085**	-0.060	
			English outcomes				
Chemistry	-0.001	-0.240***	0.074	-0.046	-0.128*	-0.019	Far (-)
English	0.028	0.592***	0.042	0.201***	0.219***	0.367***	Match (++)
Kinyarwanda	0.076	-0.028	0.101*	0.094	0.150***	-0.114	Far (-)
Math	-0.003	-0.391***	-0.042	-0.060	0.019	-0.190**	Far (-)
Gender	-0.168***	0.004	0.063	0.121**	-0.224***	0.118*	
			Kinyarwanda outcomes				
Chemistry	-0.147	-0.224***	0.030	-0.057	-0.204***	-0.176	Far (-)
English	0.049	-0.140*	0.003	-0.051	0.151*	-0.003	Far (-)
Kinyarwanda	0.255***	0.539***	0.099	0.252***	0.143***	0.265***	Match (++)
Math	0.007	-0.163*	0.024	-0.062	0.072	-0.052	Far (-)
Gender	-0.117*	-0.072	0.067	-0.062	-0.092	-0.064	
			Math outcomes				
Chemistry	0.041	0.057	0.078	0.112	-0.038	0.136*	Near (+)
English	-0.009	-0.258***	-0.043	0.057	0.057	-0.113	Far (-)
Kinyarwanda	-0.060	-0.150**	0.012	0.017	-0.100**	-0.129*	Far (-)
Math	0.505***	0.584***	0.162**	0.080	0.342***	0.403***	Match (++)
Gender	-0.043	0.143	0.111*	-0.017	-0.043	-0.116	

Note. All parameters are standardized. Within each construct, the first coefficient refers to subsample of basic-education schools (Basic). The second coefficient refers to the boarding-school subsample (Boarding). Model fit: χ^2 (df = 6878) = 7313, RMSEA = .013, CFI = .957, TLI = .955, SRMR = .072; ***p < .001; **p < .01; *p < .05.

Table 67: Path coefficients from the multiple-group regression models grouped by school type and controlling for age (manuscript two)

Predictor	Expectancies		Value		Costs		Predictions
	Day school	Boarding school	Day school	Boarding school	Day school	Boarding school	
			Chemistry outcomes				
Chemistry	0.195*	0.639***	0.081	0.293*	0.009	0.595***	Match (++)
English	-0.055	-0.088	-0.013	-0.035	0.021	-0.124	Far (-)
Kinyarwanda	0.067	-0.215***	0.104	0.043	0.019	-0.106	Far (-)
Math	0.118	-0.053	0.093	-0.081	0.151*	-0.062	Near (+)
Age	-0.037	-0.093**	0.047	-0.003	-0.069	-0.119**	
			English outcomes				
Chemistry	-0.015	-0.224***	0.085	-0.033	-0.137*	-0.027	Far (-)
English	0.030	0.586***	0.022	0.183***	0.244***	0.377***	Match (++)
Kinyarwanda	0.068	-0.039	0.102*	0.063	0.144***	-0.103	Far (-)
Math	-0.012	-0.396***	-0.036	-0.078	0.001	-0.187*	Far (-)
Age	-0.001	0.013	0.059	0.021	-0.150***	-0.044	
			Kinyarwanda outcomes				
Chemistry	-0.146*	-0.219**	0.033	-0.025	-0.208***	-0.187	Far (-)
English	0.058	-0.118*	-0.003	-0.066	0.168***	-0.022	Far (-)
Kinyarwanda	0.259***	0.526***	0.093	0.250***	0.158***	0.280***	Match (++)
Math	-0.006	-0.151*	0.033	-0.075	0.061	-0.021	Far (-)
Age	-0.093**	-0.050	0.067	0.032	-0.135***	-0.136*	
			Math outcomes				
Chemistry	0.038	0.084	0.087	0.104	-0.035	0.147**	Near (+)
English	-0.005	-0.250***	-0.052	0.083	0.057	-0.092	Far (-)
Kinyarwanda	-0.056	-0.164***	0.020	0.030	-0.088***	-0.131	Far (-)
Math	0.500***	0.605***	0.175**	0.071	0.340***	0.436***	Match (++)
Age	-0.024	0.001	0.044	0.013	-0.060	-0.067	

Note. All parameters are standardized. Within each construct, the first coefficient refers to subsample of basic-education schools (Basic). The second coefficient refers to the boarding-school subsample (Boarding). Model fit: χ^2 (df = 6878) = 7329.4, RMSEA = .013, CFI = .956, TLI = .953, SRMR = .08; ***p < .001; **p < .01; *p < .05.

Table 68: Path coefficients from the multiple-group regression models grouped by school type (S2 and S3 students only; manuscript two)

Predictor	Expectancies		Value		Costs		Predictions
	Day school	Boarding school	Day school	Boarding school	Day school	Boarding school	
			Chemistry outcomes				
Chemistry	0.203**	0.995***	0.080	0.377*	0.029	0.751***	Match (++)
English	-0.032	-0.161	0.041	-0.108	0.027	-0.155	Far (-)
Kinyarwanda	0.077	-0.312***	0.113	0.010	0.012	-0.121	Far (-)
Math	0.158	-0.062	0.114	-0.031	0.171*	-0.072	Near (+)
			English outcomes				
Chemistry	-0.032	-0.335***	0.069	-0.111	-0.139*	-0.048	Far (-)
English	0.062	0.785***	0.074	0.283***	0.270***	0.525***	Match (++)
Kinyarwanda	0.082	-0.041	0.126**	0.151	0.139**	-0.086	Far (-)
Math	0.004	-0.491***	-0.026	-0.059	0.002	-0.221*	Far (-)
			Kinyarwanda outcomes				
Chemistry	-0.170*	-0.297**	0.036	-0.059	-0.203**	-0.191	Far (-)
English	0.078	-0.179*	0.054	-0.092	0.153**	0.041	Far (-)
Kinyarwanda	0.275***	0.710***	0.113*	0.293***	0.125**	0.291***	Match (++)
Math	0.020	-0.205*	0.043	-0.047	0.057	-0.043	Far (-)
			Math outcomes				
Chemistry	0.043	0.094	0.082	0.148	-0.026	0.187*	Near (+)
English	-0.003	-0.297**	-0.012	0.046	0.020	-0.120	Far (-)
Kinyarwanda	-0.054	-0.222**	0.017	-0.007	-0.123**	-0.148*	Far (-)
Math	0.684***	0.825***	0.182**	0.076	0.384***	0.553***	Match (++)

Note. All parameters are standardized. Within each construct, the first coefficient refers to subsample of basic-education schools (Basic). The second coefficient refers to the boarding-school subsample (Boarding). Model fit: χ^2 (df = 6740) = 7157.1, RMSEA = .013, CFI = .957, TLI = .954, SRMR = .073; ***p < .001; **p < .01; *p < .05.

Table 69: Path coefficients from the multiple-group regression models grouped by school type with parallel items across domains (manuscript two)

Predictor	Expectancies		Value		Costs		Predictions
	Day school	Boarding school	Day school	Boarding school	Day school	Boarding school	
Chemistry outcomes							
Chemistry	0.147^{+}	0.664***	0.072	0.296**	0.039	0.614***	Match (++)
English	-0.049	-0.102	-0.008	-0.053	0.034	-0.150	Far (-)
Kinyarwanda	0.069	-0.211**	0.087	0.043	-0.034	-0.109**	Far (-)
Math	0.138	-0.037	0.101	-0.077	0.144*	-0.101	Near (+)
English outcomes							
Chemistry	-0.014	-0.233***	0.094	-0.063	-0.075	0.012	Far (-)
English	0.029	0.591***	0.025	0.190**	0.152***	0.411***	Match (++)
Kinyarwanda	0.069	-0.044	0.071^{+}	0.090	0.106**	-0.177**	Far (-)
Math	-0.016	-0.362***	-0.025	-0.019	-0.029	-0.201*	Far (-)
Kinyarwanda outcomes							
Chemistry	-0.141^{+}	-0.171^{+}	0.002	-0.063	-0.192**	-0.126	Far (-)
English	0.037	-0.134*	0.020	-0.034	0.139*	-0.011	Far (-)
Kinyarwanda	0.247***	0.516***	0.103	0.262***	0.132*	0.277**	Match (++)
Math	-0.008	-0.127^{+}	0.040	-0.039	0.087	-0.039	Far (-)
Math outcomes							
Chemistry	0.029	0.085	0.080	0.059	-0.029	0.207***	Near (+)
English	-0.035	-0.263**	-0.068	0.098	0.101	-0.138	Far (-)
Kinyarwanda	-0.065	-0.161**	0.052	0.069	-0.098*	-0.149**	Far (-)
Math	0.496***	0.607***	0.182**	0.132	0.320***	0.296***	Match (++)

Note. All parameters are standardized. Within each construct, the first coefficient refers to subsample of basic-education schools (Basic). The second coefficient refers to the boarding-school subsample (Boarding). Model fit: χ^2 (df = 4744) = 5040.5, RMSEA = .013, CFI = .973, TLI = .971, SRMR = .07. ***p < .001, **p < .01, *p < .05, $^{+}$p < .10.