

Students' observed engagement in lessons, instructional activities, and learning experiences

Christina Hubertina Helena Maria Heemskerk^{1,2} & Lars-Erik Malmberg¹

¹ Department of Education, Oxford University, United Kingdom

² Institut für Psychologie, Universität Bern, Switzerland

Article received 30 January 2020 / Article revised 24 July 2020 / Accepted 26 August / Available online 16 September

Abstract

In order to expand previous intraindividual studies of student engagement we investigated students' observed engagement (i.e., on- and off-task behaviour), instructional activities (i.e., teacher-led whole class, individual work, pair-work, student-teacher interaction, assessment, and "other"), and self-reported learning experiences (cognitive engagement, difficulty, competence, emotional engagement, positive and negative emotions), within lessons during one calendar week. Eighteen fourth and fifth grade target students ($M_{\text{age}}=10.1$, $SD=0.44$) were observed every 30 sec during two to four lessons each day for five school days (engagement and instructional activities), on average 66.05 times per lesson ($SD=19.16$, $\text{Range}=15-80$, $n_{\text{obs}}=14,994$) between 9-18 lessons during a week. Simultaneously, students provided 1-3 electronic questionnaire self-reports per lesson ($M_{\text{self_report}}=35.1$, $SD=12.6$, $\text{Range}=19-52$, $n_{\text{self_report}}=631$). We regressed observed engagement (0 = off-task, 1 = on-task) on self-reported learning experiences using 3-level (time-points nested in lessons, nested in students) Bayesian logistic regression models in *brms*. Observed engagement diminished during lessons, and was predicted by higher cognitive engagement, and instructional activities. As compared to teacher-led instruction, engagement was higher during individual tasks, teacher-supported tasks, and assessments. Overall self-reported and observed engagement within lessons converged, supporting their use in intraindividual research.

Keywords Intraindividual; engagement; observation; ecological momentary assessment; Bayesian



1. Introduction

Students' engagement is essential for learning. "Academic motivation refers to an individual's inclination, energy, direction, and drive with respect to learning and achievement (Martin, Ginns and Papworth, 2017). Academic engagement refers to the thoughts, behaviours, and emotions that reflect this inclination, energy, and drive (Fredricks, Blumenfeld and Paris, 2004; Martin *et al.*, 2017). Thus, engagement may be considered the manifestation of motivation—the thoughts, actions, and emotions that an individual undertakes or experiences as a result of his or her motivation (Martin, Ginns, et al., 2017)." (Collie and Martin, 2019: 2). In cross-sectional and longer-term longitudinal studies, engagement has been moderately to strongly associated with academic performance. Roorda *et al.* (2011) report a correlation of $r=0.29$ between engagement and academic achievement, whilst Godwin and Fisher (2011) report that correlations between time-on-task and learning outcomes range from 0.13 to 0.71, making student engagement an important focus of research.

To expand previous studies, we had three objectives. We first investigated how observed engagement varied *within* and between lessons, and between students. We expected engagement to vary mostly between individual students, with some situation-specific variation. In doing so, we go beyond studies in which engagement is observed during one lesson or subject (e.g. Grieco, Jowers and Bartholomew, 2009; Jarrett *et al.*, 1998), at the class level (e.g. Pöysä *et al.*, 2017; Barros, Silver and Stein, 2009), or reported on by teachers (e.g. Carlson *et al.*, 2015; Barros *et al.*, 2009) or parents (e.g. Watson *et al.*, 2019). Second, we investigated how observed engagement was related to observed instructional activities. Consistent with findings from previous studies we expected on-task behaviour to be more prevalent during one-to-one student-teacher interactions, and pair-work than during teacher-led instruction (e.g. Godwin *et al.*, 2013). Third, we investigated how observed engagement was related to self-reported learning experiences - students' thoughts, actions, and emotions reported at the beginning, middle, and end of each lesson. This goes beyond studies in which self-reports are either collected randomly throughout the school day, or systematically once a lesson. We expected within-lesson variability in on-task behaviour, and higher levels thereof, to be associated with more cognitive and emotional engagement, higher competence, less difficulty, more positive, and fewer negative emotions (e.g. Malmberg, Woolgar and Martin, 2013b).

We used the *Bayesian* estimation technique in R, implemented in *brms* (Bürkner, 2017; 2018), as this technique is well suited for rich multilevel data (14,994 time-points nested in 56 lessons nested in 18 students) with sparse observations at the highest level (Hox, Van de Schoot and Matthijsse, 2012).

1.1 Variation in observed engagement

A growing body of *intraindividual* studies focuses on situation-specific engagement (e.g. Patall *et al.*, 2016), showing large variation in engagement within students, between one situation and another. Although some studies include student-reported perceptions of the teacher (e.g., autonomy support; (Tsai *et al.*, 2008)), or global observations of classroom interaction quality (Pöysä *et al.*, 2019), there is to the best of our knowledge no previous study of students' observed intraindividual engagement during real-time instructional activities over an extended time-period. Observation studies differ in focus: either on the quality of interaction or on observable behaviours. While observation of interaction quality typically involves some degree of inference about the underlying qualities over a relatively longer sequence of time, observations of behaviours typically involve less inference and shorter but repeated observation timeframes. Momentary time sampling of behavioural engagement in young children has been found to be most accurate at short intervals (5-15 seconds) when compared to continuous recording (Zakszeski, Hojnoski and Wood, 2017). Observations of quality are typically domain-specific, focusing on agency and communion (e.g.



Wubbels *et al.*, 2015), classroom organisation (e.g., providing time to learn, minimizing disruptions), instructional support (e.g., providing explanations that support cognitive development) and emotional support (e.g., warmth, closeness) (e.g. Praetorius, Lenske and Helmke, 2012). While instructional support is associated with academic performance, and emotional support with fewer behavioural problems (La Paro and Pianta, 2000), all these dimensions are positively associated with student engagement (e.g. Malmberg *et al.*, 2010). In the present study we focused on the teachers' instructional activities, which can be considered an aspect of classroom organisation.

1.2 Observed engagement and instruction format

Students' engagement can be subdivided into behavioural, emotional, and cognitive components (Fredricks *et al.*, 2004). The key variable in the current study is behavioural engagement, as behavioural engagement can be assessed through observations in a naturalistic environment. In studies of classroom behaviour, the terms 'on-task', 'engagement', and 'attention' have been used quite interchangeably (Fredricks *et al.*, 2004). Common measures are time spent on-task or off-task (Godwin *et al.*, 2013; Pellegrini, Huberty and Jones, 1995; Pellegrini and Davis, 1993), with off-task often further categorised by type of behaviour displayed (Rock, 2005; Pellegrini and Davis, 1993), or the source of distraction (Godwin and Fisher, 2011). Following Fredricks *et al.* (2004), off-task can be grouped into active/disruptive (fidgeting, unnecessary or excessive movement) and passive/withdrawn (staring, lack of participation, dozing off) behaviours (e.g. Rock, 2005).

What constitutes on-task behaviour is inherently linked to the instructional strategy, which defines how students can carry out activities. In our study, on-task behaviour was operationalised as 'displaying goal-directed and task-appropriate behaviours', showing a clear link to the type of task set by the teacher. Whilst working in pairs, conversation with a partner is part of the activity, whereas in independent work it is not. Particularly for students in the primary school age certain task types are more conducive to high levels of on-task behaviour, such as allowing for talking and inviting collaboration. It has also been suggested that certain instruction formats are easier for teachers to supervise. Godwin *et al.* (2013) found that individual work and whole-group instruction with children seated at their desks were negatively correlated with on-task behaviour ($r=-0.018$ and $r=-0.113$ respectively), and these formats accounted for almost 37% of instruction time. In contrast, paired work was positively correlated ($r=0.032$), yet students spent only approximately 18% of their time working in pairs. Their findings are consistent with studies showing positive effects of group-work on primary school students' self-regulation (Dignath, Buettner and Langfeldt, 2008). Hence, to investigate the association between observed and experienced engagement of students, it is necessary to take account of the teacher's instructional strategy. Recently, this was confirmed by Pöysä *et al.* (2019), who found that the quality of classroom organisation was associated with students' self-reported behavioural and cognitive engagement. However, the observations were carried out in 20-minute segments at the classroom level, rather than the situational level. There is a need for studies comparing observed situation-specific behavioural engagement and self-reported engagement with learning in ordinary lessons in school.

1.3 Self-reported learning experiences

Following Schmitz (Schmitz, 2006; Schmitz and Skinner, 1993) there is a current surge in intraindividual (process) educational research due to user-friendly self-report instruments in handheld computers. Data are collected in experience sampling and ecological momentary assessment studies. There is a growing body of intraindividual studies focusing on situation-specific engagement (Martin *et al.*, 2015; Malmberg and Martin, 2019; Patall *et al.*, 2017; 2016;



Pöysä *et al.*, 2019; Shernoff *et al.*, 2016). In the present study, participants completed situation-specific ratings on their cognitive and emotional engagement, as well as competence belief, task difficulty, and emotional state. Current process approaches (Schmitz, 2006; Hamaker, 2012) emphasize the advantages of investigating processes in real-time, through multiple self-reports as these are less prone to retrospection bias and are more contextually relevant than cross sectional surveys (Wilhelm, Perrez and Pawlik, 2012). In our study students reported at the beginning, middle, and end of each lesson during one week's time, which enabled us to get snapshots of their perceptions of the task at hand (task difficulty), their current knowledge or understanding of the topic being studied (competence belief), how much thought they put into the task (cognitive engagement), whether they liked the subject or not (emotional engagement), and their emotions (positive and negative affect).

1.3.1 Self-reported engagement

School engagement is understood as a tripartite construct, its sub-components (cognitive, behavioural and emotional engagement) associated with each other (Fredricks *et al.*, 2004; Wang, Willett and Eccles, 2011). Low levels in any one of the three engagement domains has been shown to relate to unsuccessful outcomes in school, making all three – and their interaction – of interest to researchers and educators. The interplay between behavioural and cognitive engagement with learning (e.g., effort exertion, task-focus), competence beliefs (e.g., how good a student thinks he or she is at school, in a particular subject or at a task), and emotional engagement (e.g., subject liking), is posed in several theoretical models, for example, self-determination (Ryan and Deci, 2000), engagement (Fredricks *et al.*, 2004; Roorda *et al.*, 2011), and self-regulation frameworks (Boekaerts and Corno, 2005).

Cognition and cognitive activity are widely assumed to influence behaviour (Martin, 2007). It is therefore not surprising that cognitive and behavioural engagement are related constructs. Martin (2007) found that in 12,237 secondary school students (aged 12-18) adaptive cognitions and behaviours correlated ($r=0.78$), as did maladaptive cognitions and behaviours ($r=0.68$). Taking this into account, Wang *et al.* (2011) propose a model of school engagement including three second-order constructs: behavioural, emotional, and cognitive engagement. Using data from 1,103 ethnically diverse American middle-school students (8th grade), they found a correlation of $r=0.70$ between cognitive and behavioural engagement in their sample.

1.3.2 Competence beliefs and task difficulty

Although competence beliefs and task difficulty are related to each other, they are distinct. In the literature, competence belief is referred to as self-concept (Marsh, 1990; Shavelson, Hubner, & Stanton, 1976), self-efficacy (Bandura, 1997), agency beliefs (Little, 1998), and control beliefs (Skinner, Zimmer-Gembeck and Connell, 1998); converging around an individual's sense of agency (Bandura, 2008), their self-perceived capacity to fulfil a goal. For example, self-concept is positively and strongly related to academic performance (self-concept: $d=0.43$; Hattie, 2009), the association being stronger when matched within a domain (Valentine, Dubois and Cooper, 2004).

The situational equivalents of competence beliefs have been termed mastery experiences (Bandura, 1997), and competence beliefs (Malmberg *et al.*, 2013a; Tsai *et al.*, 2008). Students gauge their competence beliefs based on their history of successes and failures, providing a basis for evaluating whether subsequent tasks are deemed difficult and whether they have what it takes to succeed. Situation-specific competence belief and task difficulty have previously been found to be negatively and differentially associated (Malmberg *et al.*, 2013a). Moreover, situation-specific competence belief was positively and differentially associated with situation-specific effort exertion (Malmberg *et al.*, 2013a). On this basis, we expected higher levels of on-task behaviour to be associated with higher levels of self-reported competence, and lower levels of self-reported difficulty.



1.3.3 Subject enjoyment and learning behaviour

The enjoyment students experience in schoolwork is related to their behaviour in school. Martin (2007) found that school enjoyment correlated with adaptive academic cognitions as well as adaptive academic behaviours ($r=0.74$ and $r=0.64$ respectively). And Pietarinen, Soini and Pyhältö (2014) describe the way higher levels of emotional engagement in primary school students contribute to greater cognitive and behavioural engagement, and subsequently to higher achievement. Likewise, Hospel, Galand and Janosz (2016) identified moderate correlations between behavioural engagement - a multi-faceted construct comprising both on-task and off-task behaviours - and enjoyment, and specifically between enjoyment and classroom participation; arguably the strongest indicator of on-task behaviour ($r=0.39$ and $r=0.43$ respectively). In an earlier study, Den Brok *et al.* (2005) reported positive associations between the interpersonal closeness teachers fostered in class and their students' enjoyment of the subject taught, as well as students' self-reported effort during lessons. Finally, Wang *et al.* (2011) report a correlation of 0.89 between emotional and behavioural engagement. These results indicate that enjoyment/emotional engagement is an important factor to include when investigating behavioural engagement.

Some studies have included enjoyment as a positive emotion in momentary sampling, along with other achievement emotions (e.g. Hospel *et al.*, 2016). In this study, as observations took place during a range of subjects, enjoyment was expected to vary mostly between lessons, and have limited variability within lessons. Thus, as we postulated that enjoyment is related to - and varies between - the subjects students are studying and not so much within lessons, we separated enjoyment from the other emotions in the questionnaire, since we expected those to vary to a greater extent within lessons.

1.3.4 Emotions and learning

Emotion states in the classroom affect not only pupils' psychological well-being, but also their cognitive, motivational, and regulatory processes involved in learning and achievement (Pekrun, 2006; Goetz *et al.*, 2006). A subset of emotions, called achievement emotions, can be defined as emotions related to achievement activities or achievement outcomes (Pekrun, 2006). The circumplex model of emotions (Watson and Tellegen, 1985) categorises emotions along the dimensions of valence (pleasant vs. unpleasant) and activation (activating vs. deactivating), making it possible to distinguish four broad groups of emotions: positive activating, (e.g., enthusiasm, pride), positive deactivating (e.g., relief, relaxation, nostalgia), negative activating (e.g., anxiety, anger, and shame), and negative deactivating (e.g., boredom, hopelessness). Emotions are relevant to learning and behaviour, as they can be task-promoting or task-inhibiting (Newton, 2013; Inkinen *et al.*, 2013).

Hospel *et al.* (2016) found that different emotions relate to different types of classroom behaviour; behavioural engagement as a global measure (encompassing participation, following instructions, absenteeism, withdrawal, and disruptive behaviours) was positively related to positive emotions, and negatively to negative emotions. Each of the emotions examined in their study related to each of the five sub-domains of behavioural engagement in different ways. For example, anger related positively to all off-task behaviours (absenteeism, withdrawal, and disruptive behaviours), whereas sadness was only significantly related to absenteeism and withdrawal, but not disruptive behaviours (Hospel *et al.*, 2016). In our study we focus on on-task behaviours, rather than off-task behaviours, which can be seen as a combination of participation and following instructions. Hospel *et al.* (2016) found significant and moderate positive correlations with participation as well as following instructions for positive activating emotions (interest, hope), and weak to moderate negative correlations for negative activating emotions (anger, anxiety) and participation. Negative deactivating emotions (boredom, sadness) were weakly to moderately negatively correlated with participation. No positive deactivating emotions were included in their measures.



Unlike moods and personality traits, which are relatively much longer lasting, emotion states fluctuate throughout the day; they can last from a few seconds to a few hours (Newton, 2013). Besides being comparatively stable, traits are assumed to be mostly person-specific, whereas states are event- or situation-specific (Hamaker, Nesselroade and Molenaar, 2007). Thus, pupils' emotions at the start of a lesson may not be the same as at the end of the lesson, depending on the events and people encountered in the lesson. These fluctuations in emotion states in turn contribute to variations in (task) behaviour. Moreover, in the present study we focus on students' achievement emotions and on-task behaviour across a range of different academic domains. Goetz *et al.* (2016) found that emotions are related differently to different school subjects; for example, they found moderate positive correlations between enjoyment and English, but no significant correlation for French. Thus, achievement emotions are related to on-task behaviour, and they vary between subjects. This highlights the importance of accounting for emotional state during learning, when examining on-task behaviour in the classroom.

2. Method

2.1 Sample and procedure

Data were collected in three classrooms from three different primary schools in Southeast England (UK). A total of 62 students completed a modified version of the Learning Experience Questionnaire (Malmberg *et al.*, 2013a) 1-3 times per lesson (at the beginning, after 20 minutes, and at the end, unless the lesson ended within five minutes of the previous questionnaire). This continued for two-to-four lessons per day (between 9:50 and 16:00), for five school days (one calendar week). A sub-sample of six students per class were observed. Target students were selected by the class teacher, to provide an even split of boys and girls where possible, and include a range of attainment levels as assessed by the class teacher (under-achieving, average achievement, and high-achieving compared to age-related expectation). Mean age of the target students was 10.1 years ($SD=0.44$, range 9.2-10.6), further participant characteristics are provided in table 1. The 18 target students did not differ from the 46 non-observed students with regard to gender ($\chi^2_{[1]}=0.03$; $p=0.87$), age ($t_{[58]}=-0.93$; $p=0.36$), or within-classroom standardized academic performance ($t_{[60]}=0.51$; $p=0.61$). These target students were observed in the same order, with five seconds in between. Thus, each of the target students was observed every 30 seconds, up to a maximum of 80 observations per 40-minute lesson. The 18 students were observed 14,994 times in total, and provided 631 self-reports, in a total of 56 lessons. Observations averaged 66.05 events per lesson ($SD=19.16$, Range=15-80, $n_{obs}=14,994$), nested in an average of 12.61 lessons per student ($SD=2.75$, Range=9-18). Students provided on average 2.48 reports per lesson ($SD=0.72$, Range=1-3), on average 6.74 reports per day ($SD=2.43$, Range=2-12), totalling 35.1 reports per student ($SD=12.6$, Range=19-52). Teachers reported on students' academic performance, relative to age-related expectation, in each subject covered. Missing data occurred only for subject-specific attainment (3.1%), and in these cases generic attainment was used instead.

Ethical approval was provided by the Departmental Research Ethics Committee at the Oxford University Department for Education. Parents provided informed written consent.



Table 1

Target participant characteristics

Age		Mean 10.1; SD 0.44
		Range 9.2-10.6
Gender	Male	9 (50%)
	Female	9 (50%)
Subject-specific academic performance (% of observations linked to this level of attainment)	Well below expectation	8.3%
	Below expectation	23.6%
	At expectation	31.0%
	Above expectation	23.5%
	Well above expectation	13.6%

2.2 Measures

2.2.1 Observations

The observer (the first Author) looked up each 5 seconds to observe the engagement of each target student in turn, giving a cycle of 30 sec. After a brief observation ‘snapshot’, the type of engagement was recorded into four categories: (1) focussed on-task behaviour, (2) non-disruptive off-task behaviour, (3) off-task disruptive behaviour, and (4) absent. For the current analyses we collapsed the two off-task categories into one, giving a binary on-task variable (0=off-task, 1=on-task). ‘Absent’ ratings – when a student was not in the classroom at the time their observation turn was due – were discarded for the current analyses. Students were rated on-task if they displayed goal-directed behaviour to carrying out the task as instructed by the teacher. If their behaviour was not task-appropriate or goal-directed they were rated ‘off-task’. If they displayed goal-directed and task-appropriate actions at the same time as inappropriate actions (e.g. yawning whilst writing) and the inappropriate actions did not interfere with task completion, they were rated ‘on-task’. Where the behaviour could not definitively be appointed as on-task or off-task it was rated as ‘other’. This category also included actions such as getting a drink or going to the toilet, since these are biological needs, even though they can be used as task-avoiding behaviours. At each observation point the classroom organisation was recorded, coded into six categories (1) teacher-led whole class instruction, (2) individual student work, (3) student pair or small-group work, (4) student-teacher one-to-one interaction, (5) assessment, and (6) “other” (see Table 2).



Table 2.

Instructional activities and engagement

Instructional activities	Engagement				Total	
	Off-task		On-task		n	column %
	n	row %	n	row %		
Teacher-led whole-class	1169	24%	3725	76%	4894	33%
Individual work	1248	22%	4516	78%	5764	38%
Pair or small group	287	23%	987	77%	1274	8%
Student-teacher interaction	19	4%	516	96%	535	4%
Assessment	181	15%	987	85%	1168	8%
Other	143	11%	1216	89%	1359	9%
Total	3047	20%	11947	80%	14994	100%

Note: Six target students per class were observed every 30 sec, between two to four lessons per day, for five days.

Tasks coded as ‘teacher-led whole class instruction’ included time when pupils sat on the carpet or at desks, focused on the white board, television, or teacher at the front. Typically, they were expected to listen, process the information presented to them, or answer questions verbally/in writing. Individual tasks were those where the teacher had expressed the expectation that pupils work independently, without assistance from or discussion with other pupils. Pupils would generally be expected to focus on the materials on their own desk, along with supporting resources on the white board or wall displays. Any interaction with peers was coded as ‘off-task’. If teachers allowed collaboration, and pupils were expected to work as a pair or group to produce one collective piece of work, this was coded as pair or small group work. If pupils were expected to produce one piece of work, but copied into each individual book, this was also considered pair or small group work. If pupils were allowed to consult each other, but were expected to produce their own, unique final piece of work, this was coded as individual work, but task-related discussion with peers was coded as ‘on-task’.

Student-teacher one-to-one interaction was coded separately, due to the increased likelihood of being on-task during this interaction if it was a case of receiving additional instruction or help in task-completion. However, if a pupil was reprimanded for being off-task, or encouraged to get back on-task, and would not comply with the request, this was still coded as ‘off-task’ as instructions were not followed. Assessment was coded separate from individual work to reflect the increased pressure on pupils to work in absolute silence, and to take into account the difference in motivation between day-to-day learning tasks and test-taking. Finally, the ‘other’ category was utilised for situations where the whole class or an individual student was engaged in activities/scenarios without a specific learning outcome, yet with behavioural expectations. Examples of this are waiting for resources to be handed out, transitions between tasks, tidying the classroom, or waiting for the teacher’s assistance with a task. As shown in Table 1, students carried out individual work most of the time (38%) and interacted individually with the teacher the least (4%). Students were relatively more engaged during interaction with the teacher (96%), and



relatively less during whole-class teacher led instruction (76%). We carried out interrater agreement of the observations in another sample ($\kappa=0.80$) (Heemskerk *et al.*, 2019).

2.2.2 Self-reported learning experiences

At the beginning, middle, and end of each 40-minute lesson, students completed the brief self-report questionnaire. The questionnaire included 14 items using five-point scales. Cognitive engagement was measured with three items ("At the moment ... How much effort are you putting into completing your task?", "How focused are you on your task?" and "How much effort are you putting into keeping focused on your task?"), using the following scale: 5 = very much, 4 = quite a lot, 3 = a bit, 2 = not really, 1 = not at all (internal consistency by lesson: $M_\alpha = 0.93$, $SD_\alpha = 0.03$; and by time-point: McDonald's $\omega_{\text{time-point}} = 0.87$) (McDonald, 1999).

The following single-item measures were included, all using the same five-point scale as above: task difficulty "The task you are doing at the moment... How difficult is your task?"; competence belief "How good are you at this task?"; and subject liking "How much do you like this subject?".

Following (Watson and Tellegen, 1985) and Pekrun (2006), emotions were measured with eight items ("How are you feeling at the moment?"). Four items focused on positive emotions, two activating (i.e., alert, enthusiastic) and two deactivating (relaxed, and calm), and four on negative emotions, two activating (i.e., frustrated, angry) and two deactivating (i.e., bored, tired). The items were answered on five-point scales: 5 = very much, 4 = quite a lot, 3 = a bit, 2 = not really, 1 = not at all. Factor analysis did not support a four-factor solution, but a two-factor solution fitted reasonably. One factor represented positive emotions ($M_\alpha = 0.58$, $SD_\alpha = 0.09$; $\omega_{\text{time-point}} = 0.61$), and one negative emotions ($M_\alpha = 0.80$, $SD_\alpha = 0.02$; $\omega_{\text{time-point}} = 0.75$). Higher values for all our measures indicated more engagement, more difficulty, higher competence, and more positive and more negative emotions respectively.

Teachers reported on students' performance in relation to age-related expectations in all subjects on five-point scales (1 = well below expectation, 2 = below expectation, 3 = at age-related expectation, 4 = above expectation, 5 = well above expectation). The teacher reports were standardized within each class ($M=0$, $SD=1$). As subject-specific reports were available, we were able to link students' academic performance with each subject they did during the observed lessons, with the exception of 3.1% of observations, where a generic attainment score was used. Students' teacher-reported school-subject-specific performance is based on students' prior performance (i.e., earlier in the school year) and hence would be treated as a covariate rather than learning outcome. Academic performance was also found to be associated with students' weekly-average situation-specific ratings of competence ($r=0.46$), effort exertion ($r=0.15$) and task difficulty ($r=-0.20$) in Malmberg *et al.* (2013a).

2.2.3 Analytical procedures

We carried out all analyses using multilevel logistic regression models (Merlo *et al.*, 2006; Moineddin, Matheson and Glazier, 2007; Mood, 2010; Rozi *et al.*, 2017) in the brms R-package (Bürkner, 2017; 2018). We specified a series of three-level logistic regression models in which time-points (t) were nested within lessons (l), nested in students (s). As our focus is on associations (see Table 3) between engagement in lessons (30 sec intervals), we centred all self-reported predictors within lessons (Brincks *et al.*, 2017). R code is provided in the supplementary materials (S1).



Table 3.

Correlations between predictor variables.

	On-task	Time in lesson	Cognitive engagement (centred)	Task difficulty (centred)	Competence belief (centred)	Emotional engagement (centred)	Positive affect (centred)	Negative affect (centred)	Academic achievement (z-scored)
Time in lesson	-0.06 ***								
Cognitive engagement (centred)	0.01 ***	0.10 ***							
Task difficulty (centred)	-0.01 ***	-0.04 ***	-0.07 ***						
Competence belief (centred)	-0.03 ***	0.13 ***	0.39 ***	-0.07 ***					
Emotional engagement (centred)	-0.01 ***	0.08 ***	0.29 ***	0.17 ***	0.27 ***				
Positive affect (centred)	0.01 ***	0.01 ***	0.22 ***	0.05 ***	0.11 ***	0.09 ***			
Negative affect (centred)	0.00 ***	0.00 ***	0.08 ***	0.07 ***	0.07 ***	0.03 ***	0.29 ***		
Academic achievement (z-scored)	0.04 ***	0.04 ***	0.00 ***	-0.01 ***	0.00 ***	0.00 ***	0.00 ***	0.00 ***	
Teacher-led instruction	-0.06 ***	-0.04 ***	-0.04 ***	-0.01 ***	-0.01 ***	-0.02 **	-0.03 ***	0.00 ***	0.01 ***
Independent task	-0.03 **	0.06 ***	0.06 ***	0.00 ***	0.03 **	0.01 ***	0.04 ***	-0.02 *	0.05 ***
Paired task	-0.02 *	0.05 ***	-0.01 ***	0.02 **	0.00 ***	0.01 ***	0.01 ***	0.02 *	-0.05 ***
Teacher-student 1-2-1	0.08 ***	0.04 ***	0.01 ***	-0.03 ***	0.00 ***	-0.01 ***	0.01 ***	-0.01 ***	-0.05 ***
Test	0.03 ***	-0.12 ***	0.00 ***	0.02 *	0.02 **	0.01 ***	0.00 ***	0.02 *	-0.01 ***
Other	0.08 ***	0.01 ***	-0.03 ***	-0.01 ***	-0.04 ***	0.00 ***	-0.02 ***	0.00 ***	0.00 ***

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

We first specified a variance component model (Model 1) in which we partitioned the variance into between-lesson variance and between-student variance (Equation 1). Multilevel logistic regression models have no variance at the lowest level (Merlo *et al.*, 2006); there is no random effect for the time-points (tls). The value of the lowest level is fixed at $\frac{\pi^2}{3} = 3.29$.

$$ontask_{tls} = b_0 + v_{0s} + u_{0ls} \quad (1)$$

In Model 2 we added the within-lesson centred predictor (-1 = first 10 min, 0 = mid 20 min, 1 = last up to 10 min) of time ($time_{tls} - \overline{time_{ls}}$) (Equation 2).

$$ontask_{tls} = b_0 + b_1(time_{tls} - \overline{time_{ls}}) + v_{0s} + u_{0ls} \quad (2)$$



In Model 3 we added task-type, using teacher-led group-instruction as baseline and the other types as dummy-coded predictors: 'ind' = individual student work, 'pair' = students working in pairs or small groups, 'tea' = student-teacher one-to-one interaction, 'ass' = assessment, and 'oth' = other (Equation 3).

$$ontask_{tls} = b_0 + b_1(time_{tls} - \overline{time}_{ls}) + b_2ind_{tls} + b_3pair_{tls} + b_4tea_{tls} + b_5ass_{tls} + b_6other_{tls} + v_{0s} + u_{0ls} \quad (3)$$

In Model 4 we added self-reported learning experiences, centred within lessons: 'cogn' = cognitive engagement, 'diff' = task difficulty, 'comp' = competence belief, 'like' = emotional engagement, 'pos' = positive affect, 'neg' = negative affect (Equation 4).

$$ontask_{tls} = b_0 + b_1(time_{tls} - \overline{time}_{ls}) + b_{2-6}activities_{tls} + b_7(cogn_{tls} - \overline{cogn}_{ls}) + b_8(diff_{tls} - \overline{diff}_{ls}) + b_9(comp_{tls} - \overline{comp}_{ls}) + b_{10}(like_{tls} - \overline{like}_{ls}) + b_{11}(pos_{tls} - \overline{pos}_{ls}) + b_{12}(neg_{tls} - \overline{neg}_{ls}) + v_{0s} + u_{0ls} \quad (4)$$

In Model 5 we added teacher-reported academic performance, 'zperf', in each school-subject (Equation 5).

$$ontask_{tls} = b_0 + b_1(time_{tls} - \overline{time}_{ls}) + b_{2-6}activities_{tls} + b_{7-12}self-reports_{tls} + b_{13}zperf_{ls} + v_{0s} + u_{0ls} \quad (5)$$

The Bayesian technique estimates the probability of a parameter given the data ($p(\theta | \text{data})$), rather than the probability of the data given the model ($p(\text{data} | \text{model})$), as in null-hypothesis significance testing. It is suitable for situations when Maximum Likelihood might be underpowered, as it does not rely on large-sample theory (Zitzmann *et al.*, 2016). We specified uninformative priors for all parameters to mimic Maximum Likelihood estimates (Zitzmann *et al.*, 2016). All models were specified running four chains of 2000 iterations (of which 1000 were warm-up). Models converged well, indicated by \hat{R} values < 1.01 and efficient sample sizes were more than 100 times the number of chains (<https://mc-stan.org/misc/warnings.html>). Estimated posterior distributions and convergence of chains are presented in the supplementary materials (S2).

3. Results

3.1 Research aim 1: the variation of engagement within and between lessons, and between students.

In the baseline model, as shown in table 4, there was more variance between students than between lessons, and the most variance within lessons. In the second model we can see that observed engagement decreased during lessons ($B = -0.21$, credibility interval $[-0.30, -0.17]$, odds ratio (OR)=0.79).



3.2 Research aim 2: the relationship between observed engagement and instructional activities.

The variables entered into models 3-5 were first checked for multicollinearity. No variables caused concern as none correlated at >0.6 (Field, 2013) (see Table 3). Compared to direct instruction (teacher-led instruction), students were more engaged during individual work ($B=0.37$, $[0.26, 0.48]$, $OR=1.45$), pair-work ($B=0.29$, $[0.11, 0.48]$, $OR=1.34$), when interacting one-to-one with their teacher ($B=2.55$, $[2.10, 3.04]$, $OR=12.81$), during assessments ($B=0.36$, $[0.10, 0.61]$, $OR=1.45$), and other activities ($B=1.17$, $[0.97, 1.37]$, $OR=3.22$).

3.3 Research aim 3: the relationship between observed engagement and self-reported learning experiences.

In the fourth model, students were observed to be more behaviourally engaged when they experienced that they were more cognitively engaged than on average during the lesson ($B=0.18$, $[0.03, 0.33]$, $OR=1.20$). They were also more behaviourally engaged when they felt less competent than on average during the lesson ($B=-0.20$, $[-0.31, -0.08]$, $OR=0.83$). Task difficulty, emotional engagement, affect, and teacher-reported prior academic performance did not predict observed engagement (model 5).

For interpretation, estimated marginal means are presented in Figure 1. The distribution of on- and off-task observations across the predictor variable is visible in the black dots along the '0' and '1' values of the y-axis, as on-task is a binary variable and only can be rated 0 or 1. Each predictor has a relevant distribution of values along the x-axis, and the regression line indicates the probability (value on the y-axis) of the observation being 'on-task', given the value of the predictor on the x-axis. The shaded area around the regression line shows the credibility interval. A narrower shaded area indicates greater certainty of the estimate. Posterior distribution plots and chain plots for model 5 can be found in the supplementary materials (S2).



Table 4.

Effects of time, learning activities, learning experiences, and performance on behavioural engagement

	Model 1		Model 2		Model 3		Model 4		Model 5	
	B	s.e.	B	s.e.	B	s.e.	B	s.e.	B	s.e.
<i>Fixed effects</i>										
Intercept	1.58	0.19	1.56	0.18	1.22	0.18	1.26	0.19	1.24	0.19
Time			-0.21	0.03	-0.24	0.03	-0.24	0.03	-0.24	0.03
Individual					0.37	0.06	0.37	0.06	0.37	0.06
Pair					0.29	0.09	0.29	0.10	0.29	0.10
Teacher-student					2.56	0.25	2.55	0.25	2.55	0.24
Assessment					0.35	0.13	0.36	0.13	0.37	0.13
Other					1.17	0.10	1.16	0.10	1.17	0.10
Cognitive engagement					0.18	0.08	0.18	0.08	0.18	0.08
Task difficulty					-0.07	0.04	-0.07	0.04	-0.07	0.04
Competence belief					-0.20	0.06	-0.20	0.06	-0.19	0.06
Emotional engagement					-0.02	0.05	-0.02	0.05	-0.02	0.05
Positive emotions					0.06	0.07	0.06	0.07	0.06	0.07
Negative emotions					0.01	0.08	0.01	0.08	-0.01	0.08
Achievement									0.06	0.05
<i>Random effects</i>	σ^2		σ^2		σ^2		σ^2		σ^2	
Time-points	3.29		3.29		3.29		3.29		3.29	
Lesson ($v_{0/s}$)	0.19		0.19		0.23		0.23		0.23	
Student ($u_{0/s}$)	0.52		0.50		0.53		0.53		0.52	

Note: Posterior estimates not including zero in the credibility intervals indicated in **bold**. All estimates are from brms 2.10.0 (2019-08-27)

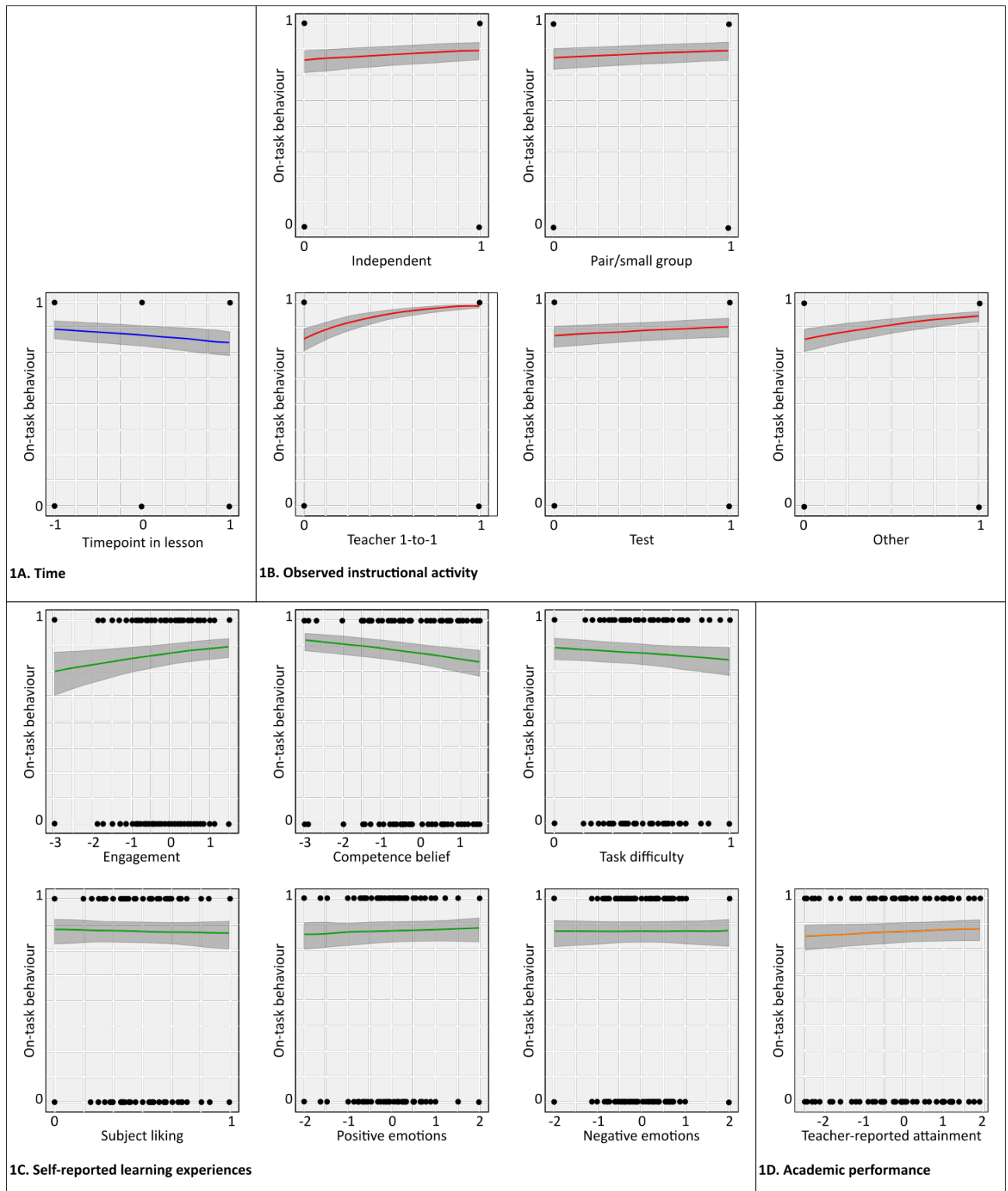


Figure 1.

Estimated marginal means for fixed effects. All estimates are from brms 2.10.0 (2019-08-27)

Note: values on the y-axis represent the probability of the outcome variable (behavioral engagement, 0 = off-task, 1 = on-task). Predictor effects (x-axis) follow the metric of the predictor. 1A:



time. 0 = beginning of lesson, 1 = mid-lesson, 2 = end-of-lesson. 1B: task type. Individual work, pair-work, teacher-student interaction, assessment, and other (0 = not observed, 1 = observed). 1C: self-reported learning experiences. Cognitive engagement, task difficulty, competence belief, emotional engagement, positive affect, negative affect (within-lesson centered scores, original scale 1-5). 1D: within-classroom z-scored subject specific academic performance.

4. Discussion

Using Bayesian estimation, we investigated how observed engagement (1) varied *within* and between lessons, and between students, (2) was associated with instructional activities and (3) was associated with self-reported learning experiences (controlling for academic performance). Most variance was found within lessons, between students, and least between lessons. Observed student engagement decreased from the beginning to the end of lessons, and associated with the type of learning activity in the class (one-to-one student-teacher interaction > “other” activities > individual work > assessment > pair/group work > teacher-led whole class instruction). Observed engagement was higher when students reported higher cognitive engagement and lower competence belief.

4.1 Variation in observed engagement within and between lessons, and between students

Investigating situation-specific motivation in schools, Martin *et al.* (2015) identified substantial variation in student engagement within days and between students. In line with their findings, we found considerable variation at the situation-level (within lessons in our study), and between students. Direct comparison of the level of variation is not meaningful, as our data structure diverged by further differentiating within-lesson and between-lesson observations and self-reports, whilst not including day-to-day variation. However, whereas Martin *et al.* found the greatest degree of variation at the between-student level, we found greater situation-specific variation. Our study might uniquely suggest that the more fine-tuned the manner in which we observe behaviour and ask for self-reports, the more variance we find at this level (30 second intervals for observations, 15-20 minutes for self-reports). The greater situational variance in on-task behaviour in our sample may be related to the age of the participants; our sample consisted of primary rather than secondary school students, and it is suggested that with increased age and metacognitive strategy use, the dependence of engagement on contextual factors decreases (Fredricks *et al.*, 2004).

4.2 Instruction format and observed engagement

In their study of instruction format in relation to on-task behaviour in four- to eight-year-old children, Godwin *et al.* (2013) found that students were least on-task during individual work and whole-class instruction, and most on-task during small-group work and testing. Their categories did not include ‘one-to-one support’ or ‘other’. Our results also indicated higher levels of on-task behaviour during testing, and lower levels during whole-class instruction. However, we found that students were more on-task during individual work, and less during pair/group work. This may relate to the age of participants, as our sample was older than in the study by Godwin *et al.*, and they might be better able to cope with working independently. Godwin *et al.* (2013) suggest that certain instructional formats present greater possibility for students go off-task, as they are



more difficult for teachers to supervise. This would be a logical explanation for the relatively high level of on-task behaviour in one-to-one instructional settings in the present study.

In the present study, 33% of observations involved teacher-led, whole-class instruction, and 38% involved independent work. Only 8% of the observed time was spent working in pairs or small groups. Due to the small proportion of time spent working in small groups, our results for this category of task should be interpreted with caution. Similar to the findings of Godwin *et al.* (2013), overall associations between observed engagement (on-task) and instructional activities were small.

4.3 Self-reported learning experiences and observed engagement

Within-lesson increases in cognitive engagement were expected to be positively related to observed behavioural engagement, and this was confirmed by the data. Competence belief unexpectedly related negatively to behavioural engagement. This may be an indication that pupils felt a greater need to pay attention and participate in classroom activities at times when they perceived their competence at the task in hand to be lower than average. We found no significant relationship between on-task behaviour and within-lesson changes in task difficulty, emotional engagement, or emotion states. As expected, emotional engagement may vary more strongly between lessons than within lesson. Emotion states may have been too difficult for children of this age to accurately identify and rate, despite efforts to only include emotions in the questionnaire that children in this age category would be familiar with. Another possibility is that the differential effects of activating and deactivating emotions of the same valence may be masking one another by including only two emotion state factors (positive and negative affect) in our analyses. However, factor analysis did not support a four-factor structure as suggested in the circumplex model (positive activating, positive deactivating, negative activating, negative deactivating (Watson and Tellegen, 1985)) in our data, but rather a two-factor structure (positive and negative affect). Due to the repeated measurements taken, it was decided to keep the questionnaire as short as possible. This, in combination with the limited vocabulary of the age group for naming emotions, led to only two emotions being included for each quadrant of the circumplex model. It may be that, with three or four emotions per quadrant, a four-factor model would have been supported by the data, and these in turn might have been related to behavioural engagement in this age group. Overall associations between observed engagement (on-task) and self-reported learning experiences were small.

4.4 Limitations

The present study had some limitations. First, while we had rich data at the situational level ($n = 14,994$) the sample size at the student-level was small ($n = 18$). Although three diverse schools were included in the current sample (state-funded, voluntary-aided, and private), further sociodemographic data for participants was not collected. Further replications of our findings in larger samples are needed, including the collection of individual sociodemographic data. Second, while we had rich data on instructional activity, we did not investigate the quality of the student-teacher interaction during the instructional activity. Such future studies could show whether quality of student-teacher interaction varies across instructional activities. For example, emotional support could be higher in one-to-one student-teacher interaction (La Paro and Pianta, 2000). Third, while we did not carry out interrater agreement assessments with the current version of the observation instrument, we did so in Heemskerk *et al.* (2019). As cognitive engagement cannot be measured by observations, we relied on self-reports for this. Finally, while we focussed on exploration of variance within lessons, further analyses could focus on differences between lessons.



4.5 Conclusion

Observed engagement varied more greatly within lessons than between lessons, and diminished over time during lessons. On-task behaviour was predicted by higher cognitive engagement and instructional activities. As compared to teacher-led instruction, engagement was higher during individual tasks, teacher-supported tasks, and assessments. Overall self-reported (cognitive) and observed (behavioural) engagement within lessons converged, supporting their use in intraindividual research.

Keypoints

- Student engagement varies substantially within lessons and between students.
- More variance in behaviour is found at the situational level with more fine-tuned observation and self-report methods.
- Student engagement is related to instructional activities, with the highest engagement during one-to-one instruction, assessments, and independent work.
- Educators must carefully consider their use of whole-group instructions, as this accounted for 33% of observed time and was associated with the lowest level of engagement.
- Observed engagement is positively related to self-reported engagement and negatively related to competence belief.

References

- Bandura, A. (1997) *Self-efficacy: the exercise of control*, New York: W.H. Freeman.
- Bandura, A. (2008) Toward an agentic theory of the self. In: Marsh, H.W., Craven, R.G. and McInerney, D.M. (eds) *Self-processes, learning, and enabling human potential: dynamic new approaches*. Charlotte, NC: Information Age Publishing, pp.15-49.
- Barros, R.M., Silver, E.J. and Stein, R.E. (2009) 'School recess and group classroom behavior'. *Pediatrics* 123(2) pp.431-436. DOI: 10.1542/peds.2007-2825.
- Boekaerts, M. and Corno, L. (2005) 'Self-Regulation in the Classroom: A Perspective on Assessment and Intervention'. *Applied Psychology* 54(2) pp.199-231. DOI: 10.1111/j.1464-0597.2005.00205.x.
- Brincks, A.M., Enders, C.K., Llabre, M.M., Bulotsky-Shearer, R.J., Prado, G. and Feaster, D.J. (2017) 'Centering Predictor Variables in Three-Level Contextual Models'. *Multivariate Behavioral Research* 52(2) pp.149-163. DOI: 10.1080/00273171.2016.1256753.
- Bürkner, P.-C. (2017) 'brms: An R Package for Bayesian Multilevel Models Using Stan'. *Journal of Statistical Software* 80(1) pp.1-28. DOI: 10.18637/jss.v080.i01.
- Bürkner, P.-C. (2018) 'Advanced Bayesian Multilevel Modeling with the R Package brms'. *The R Journal* 10(1). DOI: 10.32614/RJ-2018-017.
- Carlson, J.A., Engelberg, J.K., Cain, K.L., Conway, T.L., Mignano, A.M., Bonilla, E.A., Geremia, C. and Sallis, J.F. (2015) 'Implementing classroom physical activity breaks: Associations with student physical activity and classroom behavior'. *Preventive medicine* 81 pp.67-72. DOI: 10.1016/j.ypmed.2015.08.006.
- Collie, R.J. and Martin, A.J. (2019) Motivation and engagement in learning. *Oxford Research Encyclopedia of Education*. New York: Oxford University Press.
- Den Brok, P., Levy, J., Brekelmans, M. and Wubbels, T. (2005) 'The effect of teacher interpersonal behaviour on students' subject-specific motivation'. *The Journal of Classroom Interaction* 40.



- Dignath, C., Buettner, G. and Langfeldt, H.-P. (2008) 'How can primary school students learn self-regulated learning strategies most effectively?'. *Educational Research Review* 3(2) pp.101-129. DOI: 10.1016/j.edurev.2008.02.003.
- Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*, London: SAGE.
- Fredricks, J.A., Blumenfeld, P.C. and Paris, A.H. (2004) 'School Engagement: Potential of the Concept, State of the Evidence'. *Review of Educational Research* 74(1) pp.59-109. DOI: 10.3102/00346543074001059.
- Godwin, K.E., Almeda, M., Petroccia, M., Baker, R.S. and Fisher, A.V. (2013) 'Classroom activities and off-task behavior in elementary school children', *35th Annual Conference of the Cognitive Science Society*. Austin, Tx, Cognitive Science Society. pp.2428-2433.
- Godwin, K.E. and Fisher, A.V. (2011) 'Allocation of Attention in Classroom Environments: Consequences for Learning', *33rd Annual Conference of the Cognitive Science Society*. Austin, Tx, Cognitive Science Society. pp.2806-2811.
- Goetz, T., Pekrun, R., Hall, N. and Haag, L. (2006) 'Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction'. *British Journal of Educational Psychology* 76 pp.289-308. DOI: 10.1348/000709905X42860.
- Goetz, T., Sticca, F., Pekrun, R., Murayama, K. and Elliot, A.J. (2016) 'Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach'. *Learning and Instruction* 41(C) pp.115-125. DOI: 10.1016/j.learninstruc.2015.10.007.
- Grieco, L.A., Jowers, E.M. and Bartholomew, J.B. (2009) 'Physically active academic lessons and time on task: the moderating effect of body mass index'. *Medicine and Science in Sports and Exercise* 41(10) pp.1921-1926. DOI: 10.1249/MSS.0b013e3181a61495.
- Hamaker, E.L. (2012) Why researchers should think "within-person": A pragmatic rationale. In: Mehl, M.R. and Conner, T.S. (eds) *Handbook of research methods for studying daily life*. London: Guilford Press, pp.43-61.
- Hamaker, E.L., Nesselroade, J.R. and Molenaar, P.C.M. (2007) 'The integrated trait-state model'. *Journal of research in personality* 41(2) pp.295-315. DOI: 10.1016/j.jrp.2006.04.003.
- Hattie, J. (2009) *Visible Learning*, Abingdon: Routledge.
- Hospel, V., Galand, B. and Janosz, M. (2016) 'Multidimensionality of behavioural engagement: Empirical support and implications'. *International Journal of Educational Research* 77 pp.37-49. DOI: <https://doi.org/10.1016/j.ijer.2016.02.007>.
- Hox, J.J., Van de Schoot, R. and Matthijsse, S. (2012) 'How few countries will do? Comparative survey analysis from a Bayesian perspective'. *Survey Research Methods* 6(2) pp.87. DOI: 10.18148/srm/2012.v6i2.5033.
- Inkinen, M., Lonka, K., Hakkarainen, K., Muukkonen, H., Litmanen, T. and Salmela-Aro, K. (2013) 'The Interface Between Core Affects and the Challenge– Skill Relationship'. *Journal of Happiness Studies* 15 pp.891-913. DOI: 10.1007/s10902-013-9455-6.
- Jarrett, O.S., Maxwell, D.M., Dickerson, C., Hoge, P., Davies, G. and Yetley, A. (1998) 'Impact of Recess on Classroom Behavior: Group Effects and Individual Differences'. *The Journal of Educational Research* 92(2) pp.121-126. DOI: 10.1080/00220679809597584.
- La Paro, K.M. and Pianta, R.C. (2000) 'Predicting Children's Competence in the Early School Years: A Meta-Analytic Review'. *Review of Educational Research* 70(4) pp.443-484. DOI: 10.3102/00346543070004443.
- Little, T.D. (1998) Sociocultural influences on the development of children's action-control beliefs. In: Heckhausen, J. and Dweck, C.S. (eds) *Motivation and self-regulation across the life span*. Cambridge, USA: Cambridge University Press, pp.281-315.
- Malmberg, L.-E., Hagger, H., Burn, K., Mutton, T. and Colls, H. (2010) 'Observed classroom quality during teacher education and two years of professional practice'. *Journal of Educational Psychology* 102(4) pp.916-932. DOI: 10.1037/a0020920.
- Malmberg, L.-E. and Martin, A.J. (2019) 'Processes of students' effort exertion, competence beliefs and motivation: Cyclic and dynamic effects of learning experiences within school days and school



- subjects'. *Contemporary Educational Psychology* 58 pp.299-309. DOI: 10.1016/j.cedpsych.2019.03.013.
- Malmberg, L.-E., Walls, T.A., Martin, A.J., Little, T.D. and Lim, W.H.T. (2013a) 'Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models'. *Learning and Individual Differences* 28 pp.54-65. DOI: <https://doi.org/10.1016/j.lindif.2013.09.007>.
- Malmberg, L.-E., Woolgar, C. and Martin, A.J. (2013b) 'Quality of measurement of the learning experience questionnaire for personal digital assistants'. *International journal of quantitative research in education* 1(3) pp.275-296.
- Martin, A.J. (2007) 'Examining a multidimensional model of student motivation and engagement using a construct validation approach'. *The British journal of educational psychology* 77(2) pp.413-440. DOI: 10.1348/000709906X118036.
- Martin, A.J., Ginns, P. and Papworth, B. (2017) 'Motivation and engagement: Same or different? Does it matter?'. *Learning and Individual Differences* 55 pp.150-162. DOI: 10.1016/j.lindif.2017.03.013.
- Martin, A.J., Papworth, B., Ginns, P., Malmberg, L.-E., Collie, R.J. and 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(Supplement C) pp.26-35. DOI: <https://doi.org/10.1016/j.lindif.2015.01.014>.
- McDonald, R.P. (1999) *Test theory: A unified treatment*, Hillsdale: Erlbaum.
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L. and Larsen, K. (2006) 'A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena'. *Journal of Epidemiology and Community Health* 60(4) pp.290. DOI: 10.1136/jech.2004.029454.
- Moineddin, R., Matheson, F. and Glazier, R.H. (2007) 'A simulation study of sample size for multilevel logistic regression models'. *BMC Med. Res. Methodol.* 7(1). DOI: 10.1186/1471-2288-7-34.
- Mood, C. (2010) 'Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It'. *European Sociological Review* 26(1) pp.67-82. DOI: 10.1093/esr/jcp006.
- Newton, D.P. (2013) 'Moods, emotions and creative thinking: A framework for teaching'. *Thinking Skills and Creativity* 8(Supplement C) pp.34-44. DOI: <https://doi.org/10.1016/j.tsc.2012.05.006>.
- Patall, E.A., Vasquez, A.C., Steingut, R.R., Trimble, S.S. and Pituch, K.A. (2016) 'Daily interest, engagement, and autonomy support in the high school science classroom'. *Contemporary Educational Psychology* 46 pp.180-194. DOI: 10.1016/j.cedpsych.2016.06.002.
- Patall, E.A., Vasquez, A.C., Steingut, R.R., Trimble, S.S. and Pituch, K.A. (2017) 'Supporting and Thwarting Autonomy in the High School Science Classroom'. *Cognition and Instruction* 35(4) pp.337-362. DOI: 10.1080/07370008.2017.1358722.
- Pekrun, R. (2006) 'The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice'. *Educational Psychology Review* 18(4) pp.315-341. DOI: 10.1007/s10648-006-9029-9.
- Pellegrini, A.D. and Davis, P. (1993) 'Relations between children's playground and classroom behaviour'. *British Journal of Educational Psychology* 63(1) pp.88-95. DOI: 10.1111/j.2044-8279.1993.tb01043.x.
- Pellegrini, A.D., Huberty, P.D. and Jones, I. (1995) 'The Effects of Recess Timing on Children's Playground and Classroom Behaviors'. *American Educational Research Journal* 32(4) pp.845-864. DOI: 10.3102/00028312032004845.
- Pietarinen, J., Soini, T. and Pyhäntö, K. (2014) 'Students' emotional and cognitive engagement as the determinants of well-being and achievement in school'. *International Journal of Educational Research* 67 pp.40-51. DOI: <https://doi.org/10.1016/j.ijer.2014.05.001>.
- Pöysä, S., Vasalampi, K., Muotka, J., Lerkkanen, M.-K., Poikkeus, A.-M. and Nurmi, J.-E. (2017) 'Variation in situation-specific engagement among lower secondary school students'. *Learning and Instruction* 53(Supplement C) pp.64-73. DOI: <https://doi.org/10.1016/j.learninstruc.2017.07.007>.



- Pöysä, S., Vasalampi, K., Muotka, J., Lerkkanen, M.K., Poikkeus, A.M. and Nurmi, J.E. (2019) 'Teacher–student interaction and lower secondary school students' situational engagement'. *British Journal of Educational Psychology* 89(2) pp.374-392. DOI: 10.1111/bjep.12244.
- Praetorius, A.-K., Lenske, G. and Helmke, A. (2012) 'Observer ratings of instructional quality: Do they fulfill what they promise?'. *Learning and Instruction* 22(6) pp.387-400. DOI: 10.1016/j.learninstruc.2012.03.002.
- Rock, M.L. (2005) 'Use of Strategic Self-Monitoring to Enhance Academic Engagement, Productivity, and Accuracy of Students With and Without Exceptionalities'. *Journal of Positive Behavior Interventions* 7(1) pp.3-17. DOI: 10.1177/10983007050070010201.
- Roorda, D.L., Koomen, H.M.Y., Spilt, J.L. and Oort, F.J. (2011) 'The Influence of Affective Teacher–Student Relationships on Students' School Engagement and Achievement'. *Review of Educational Research* 81(4) pp.493-529. DOI: 10.3102/0034654311421793.
- Rozi, S., Mahmud, S., Lancaster, G., Hadden, W. and Pappas, G. (2017) 'Multilevel Modeling of Binary Outcomes with Three-Level Complex Health Survey Data'. *Open Journal of Epidemiology* 7(1) pp.17. DOI: 10.4236/ojepi.2017.71004.
- Ryan, R. and Deci, E. (2000) 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being'. *American Psychologist* 55(1) pp.68-78. DOI: 10.1037/0003-066x.55.1.68.
- Schmitz, B. (2006) 'Advantages of studying processes in educational research'. *Learning and Instruction* 16(5) pp.433-449. DOI: 10.1016/j.learninstruc.2006.09.004.
- Schmitz, B. and Skinner, E. (1993) 'Perceived Control, Effort, and Academic Performance: Interindividual, Intraindividual, and Multivariate Time-Series Analyses'. *Journal of Personality and Social Psychology* 64(6) pp.1010. DOI: 10.1037/0022-3514.64.6.1010.
- Shernoff, D.J., Kelly, S., Tonks, S.M., Anderson, B., Cavanagh, R.F., Sinha, S. and Abdi, B. (2016) 'Student engagement as a function of environmental complexity in high school classrooms'. *Learning and Instruction* 43(C) pp.52-60. DOI: 10.1016/j.learninstruc.2015.12.003.
- Skinner, E.A., Zimmer-Gembeck, M.J. and Connell, J.P. (1998) 'Individual differences and the development of perceived control - Introduction and overview'. *Monographs of the Society for Research in Child Development* 63(254). DOI: 10.2307/1166220.
- Tsai, Y.-M., Kunter, M., Lüdtke, O. and Trautwein, U. (2008) Day-to-day variation in competence beliefs: How autonomy support predicts young adolescents' felt competence. In: Marsh, H.W., Craven, R.G. and McInerney, D.M. (eds) *Self-processes, learning, and enabling human potential: dynamic new approaches*. Charlotte, NC: Information Age Pub., pp.119-120.
- Valentine, J.C., Dubois, D.L. and Cooper, H. (2004) 'The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review'. *Educational Psychologist* 39(2) pp.111-133. DOI: 10.1207/s15326985ep3902_3.
- Wang, M.-T., Willett, J.B. and Eccles, J.S. (2011) 'The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity'. *Journal of School Psychology* 49(4) pp.465-480. DOI: 10.1016/j.jsp.2011.04.001.
- Watson, A., Timperio, A., Brown, H., Hinkley, T. and Hesketh, K.D. (2019) 'Associations between organised sport participation and classroom behaviour outcomes among primary school-aged children'. *PLoS One* 14(1). DOI: 10.1371/journal.pone.0209354.
- Watson, D. and Tellegen, A. (1985) 'Toward a consensual structure of mood'. *Psychological Bulletin* 98(2) pp.219. DOI: 10.1037/0033-2909.98.2.219.
- Wilhelm, P., Perez, M. and Pawlik, K. (2012) Conducting research in daily life. In: Mehl, M.R. and Conner, T.S. (eds) *Handbook of research methods for studying daily life*. London: Guilford Press, pp.62-86.
- Wubbels, T., Brekelmans, M., Den Brok, P., Wijsman, L., Mainhard, T. and Van Tartwijk, J. (2015) Teacher-student relationships and classroom management. In: Emmer, E.T. and Sabornie, E.J. (eds) *Handbook of classroom management*. Second edition ed. New York: Routledge, pp.363-386.



- Zakszeski, B.N., Hojnoski, R.L. and Wood, B.K. (2017) 'Considerations for Time Sampling Interval Durations in the Measurement of Young Children's Classroom Engagement'. *Topics In Early Childhood Special Education* 37(1) pp.42-53. DOI: 10.1177/0271121416659054.
- Zitzmann, S., Lüdtke, O., Robitzsch, A. and Marsh, H.W. (2016) 'A Bayesian Approach for Estimating Multilevel Latent Contextual Models'. *Structural Equation Modeling: A Multidisciplinary Journal* 23(5) pp.661-679. DOI: 10.1080/10705511.2016.1207179.