An international perspective on Teacher's self-efficacy: Investigating personal, cultural, structural and environmental factors

---

Sina Veronika Fackler
Department of Education
University of Oxford
Kellogg College



Supervisors
Professor Lars-Erik Malmberg
Professor Pam Sammons

Thesis submitted to the University of Oxford in fulfilment of the requirements for the degree of Doctor of Philosophy in Education
Trinity Term, 2018

ABSTRACT

This study investigates the variation of predictors of teachers' self-efficacy (TSE) in 32 OECD countries and a sub-set of three country clusters containing Anglo-Saxon, Nordic and East and South-East Asian countries (ESEA), offering a cultural perspective on TSE. TSE is defined as a tripartite measure comprising TSE in student engagement, instruction and classroom management. It has been found to be a useful tool to explain differences as to why some teachers are convinced of their ability to bring about change in their students' learning, while others do not believe in their abilities. The study makes use of the Teaching and Learning International Survey (TALIS) 2013, published by the OECD, that has information of about 104.000 teachers in about 6.500 schools. It is one of very few multi-level studies in the field and brings together variables describing the teachers, the classroom, the principal and the school.

Firstly, it was found that the three domains of TSE are related to the predictors in different ways and most variation in terms of the relationship was found for the classroom setting. Proportionally, most variation could be explained within schools, whereas individual differences between teachers accounted for most variation in TSE within teachers. Secondly, the predictors were associated differently in each domain of TSE across the three country clusters, as i.e. most classroom predictors were significant in the ESEA countries, but rather few in the Nordic countries, whereas student background features were relevant for explaining variation in the Anglo-Saxon countries.

This study was able to contribute to providing levers for education policy and practice to improve the workplace of teachers. It shed light on to how teachers in different cultural settings experience TSE in different ways, although they work in similar institutions. The large sample size allowed robust and generalizable results as well as the application of advanced statistical methods like multilevel- and structure equation modelling. Hence, the study contributes to knowledge on a methodological, practical and theoretical level.

ACKNOWLEDGEMENTS

First of all, I would like to express my profound gratitude to my supervisors, Professor Lars-Erik Malmberg and Professor Pam Sammons, for their invaluable support and guidance during my DPhil research. Their sound advice, experience, ideas and knowledge as well as teaching of the literature and state-of-the-art complex statistical techniques, but also their personal support made my doctoral studies a stimulating and productive experience and contributed to the completion of my thesis.

My sincere gratitude goes to the ESRC Quantitative Methods Scholarship Scheme for their financial support, which enabled me to take part in a stimulating and vibrant international academic environment, to benefit from high-stakes conferences and to take part in training programmes. I would also like to thank Kellogg College, University Oxford, and the Department of Education, University of Oxford, for providing a supportive academic and social setting.

For their personal support I would like to thank friends and colleagues from college, the department and the catholic community in Oxford, who have accompanied me.

Finally, I cannot thank my husband Martin and my family in Germany enough for all their love, support and encouragement they have provided me with during my doctoral research.

TABLE OF CONTENTS

ABSTRACT **2**

ACKNOWLEDGEMENTS **3**

DISSERTATION OVERVIEW **17**

CHAPTER 1 INTRODUCTION **20**

1.1 RATIONALE FOR THE RESEARCH **20**

1.2 SCOPE OF THE RESEARCH **22**

CHAPTER 2 THEORETICAL FRAMEWORK AND LITERATURE REVIEW **25**

2.1 THEORIES LEADING TO TSE **25**

2.1.1 SOCIAL LEARNING THEORY 25

2.1.2 SOCIAL COGNITIVE THEORY 26

2.1.3 ATTRIBUTION THEORY 28

2.1.4 SUMMARY 30

2.2 INDIVIDUAL CHARACTERISTICS OF TSE **30**

2.2.1 EXAMPLE OF TWO MEASUREMENTS USING ROTTER'S 'LOCUS OF CONTROL' 31

2.2.1.1 The RAND-measurement 31

2.2.1.2 Guskey's concept of 'Responsibility of student achievement' 32

2.2.1.3 Rose & Medway's contextualisation of TSE 32

2.2.2 EXAMPLES OF MEASUREMENTS USING BANDURA'S CONCEPT OF SELF-EFFICACY 33

2.2.2.1 Gibson & Dembo's combination of the RAND-measure and Bandura's Outcome expectancies 33

2.2.2.2 Bandura's conceptualisation of a measurement of TSE 34

2.2.2.3 Teacher self-efficacy scale by Schwarzer & Hallum, 2008 35

2.2.2.4 Teacher self-efficacy scale by Skaalvik & Skaalvik, 2017 36

2.2.2.5 *The Ohio State Teacher Self-efficacy scale (OSTES)* 37

2.2.3 SUMMARY 38

2.3 PAST INVESTIGATIONS OF TEACHER'S SELF-EFFICACY **39**

2.3.1 TEACHER CHARACTERISTICS AND TEACHER'S SELF-EFFICACY 39

2.3.2 Summary 41
2.3.3 Classroom characteristics and teacher's self-efficacy 41
2.3.4 Summary 44
2.3.5 Principal and school characteristics and teacher's self-efficacy 44
2.3.6 Summary 46
2.3.7 Country characteristics of and cultural values related to TSE 47

CHAPTER 3 METHODOLOGY 49

3.1 Ethical considerations 49
3.2 Secondary data analysis 50
3.2.1 Epistemological foundations of quantitative secondary data analysis 50
3.2.2 Advantages of Secondary data analysis 51
3.2.3 Disadvantages of secondary data analysis 52
3.3 The OECD, education and TALIS 53
3.3.1 The OECD and education 53
3.3.2 The TALIS study 54
3.3.3 Compilation of the questionnaire 54
3.3.4 Sampling 55
3.3.5 Data collection 56
3.3.6 Implications for the use of OECD data 57
3.3.7 Summary 59
3.4 Analytic procedure 59
3.4.1 Multilevel analysis 59
3.4.2 Sources of error 62
3.4.3 Types of variables in multi-level data 62
3.4.4 Centering variables 63
3.4.5 The intra-class correlation 64
3.4.6 Analytical procedure: How is TSE in TALIS 2013 predicted by teacher characteristics, classroom characteristics and characteristics of the school and principal? 65
3.5 Description of variables 68
3.5.1 Teacher level (Teacher and classroom characteristics) 69
3.5.1.1 Teacher characteristics 69
3.5.1.2 Classroom characteristics 69

3.5.2 PRINCIPAL LEVEL (PRINCIPAL AND SCHOOL CHARACTERISTICS) 70
3.5.2.1 Principal characteristics 70
3.5.2.2 School characteristics 71
3.6 QUALITY OF THE SCALES USED 71
3.6.1 MULTILEVEL STRUCTURE EQUATION MODELLING 71
3.6.2 MEASUREMENT INVARIANCE TESTING 73
3.6.3 MEASUREMENT OF THE DEPENDENT VARIABLE: TEACHER SELF-EFFICACY 75
3.6.4 INSTRUCTIONAL STYLE OF LEADERSHIP 81
3.6.5 CLASSROOM CLIMATE 84
3.6.6 CONSTRUCTIVIST BELIEFS ABOUT TEACHING 87
3.6.7 SUMMARY 90
3.7 MISSING DATA 91
3.7.1 MISSING DATA MECHANISMS 91
3.7.2 MISSING COMPLETELY AT RANDOM (MCAR) 92
3.7.3 MISSING AT RANDOM (MAR) 92
3.7.4 MISSING NOT AT RANDOM (MNAR) 93
3.8 APPROACHES TO HANDLING MISSING DATA 94
3.8.1 DELETION METHODS: COMPLETE AND AVAILABLE CASE ANALYSIS 94
3.8.1.1 List wise deletion 95
3.8.1.2 Pairwise deletion 95
3.8.2 SINGLE IMPUTATION TECHNIQUES 96
3.8.2.1 Mean imputation 96
3.8.2.2 Regression imputation and stochastic regression imputation 96
3.8.2.3 Hot-deck imputation 97
3.8.3 FULL INFORMATION MAXIMUM LIKELIHOOD IMPUTATION (FIML) AND MULTIPLE IMPUTATION (MI) 98
3.8.3.1 Full information maximum likelihood imputation 98
3.8.3.2 Multiple imputation 99
3.8.3.3 MI and Multilevel data 101
3.9 IMPUTATION OF MISSING VALUES 102
3.10 MISSING VALUE PATTERNS AND STATISTICS 105
3.10.1 TEACHER VARIABLES 105
3.10.2 CLASSROOM VARIABLES 108
3.10.3 PRINCIPAL VARIABLES 110
3.10.4 SCHOOL VARIABLES 112

3.11 SUMMARY 114

CHAPTER 4 RESULTS: PREDICTING TEACHERS' SELF-EFFICACY BY TEACHER-, CLASSROOM-, PRINCIPAL- AND SCHOOL CHARACTERISTICS 115

4.1 TEACHER'S SELF-EFFICACY IN STUDENT ENGAGEMENT (ETSE) 119
4.1.1 BASELINE MODEL 119
4.1.2 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY TEACHER CHARACTERISTICS? 119
4.1.3 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY CLASSROOM CHARACTERISTICS? 121
4.1.4 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY PRINCIPAL CHARACTERISTICS? 124
4.1.5 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY SCHOOL CHARACTERISTICS? 125
4.2 TEACHERS' SELF-EFFICACY IN INSTRUCTION (ITSE) 126
4.2.1 BASELINE MODEL 126
4.2.2 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY TEACHER CHARACTERISTICS? 126
4.2.3 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY CLASSROOM CHARACTERISTICS? 128
4.2.4 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY PRINCIPAL CHARACTERISTICS? 130
4.2.5 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY SCHOOL CHARACTERISTICS? 131
4.3 TEACHERS' SELF-EFFICACY IN CLASSROOM MANAGEMENT (MTSE) 132
4.3.1 BASELINE MODEL 132
4.3.2 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY TEACHER CHARACTERISTICS? 132
4.3.3 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY CLASSROOM CHARACTERISTICS? 135
4.3.4 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY PRINCIPAL CHARACTERISTICS? 136
4.3.5 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY SCHOOL CHARACTERISTICS? 137
4.4 COMPARATIVE MODEL SUMMARIES 138
4.4.1 BASELINE MODEL SUMMARY 138
4.4.2 1ST MODEL COMPARATIVE SUMMARY 139
4.4.3 2ND MODEL COMPARATIVE SUMMARY 141
4.4.4 3RD MODEL COMPARATIVE SUMMARY 143
4.4.5 4TH MODEL COMPARATIVE SUMMARY 144
4.5 SUMMARY 145

CHAPTER 5 COUNTRY-SPECIFIC CHARACTERISTICS OF SCHOOL SYSTEMS AND SOCIETIES 148

5.1 CATEGORISATION OF EDUCATION SYSTEMS 149

5.2 SUMMARY 152
5.3 COUNTRY VALUES AND LEADERSHIP 153
5.3.1 POWER DISTANCE 154
5.3.2 GENDER EGALITARIANISM 155
5.3.3 COLLECTIVISM AND INDIVIDUALISM 156
5.3.4 PERFORMANCE ORIENTATION 157
5.3.5 SUMMARY 158
5.4 TEACHING STYLES AND CULTURE 158
5.4.1 TEACHING AND EDUCATION IN ANGLO-SAXON COUNTRIES 158
5.4.2 SUMMARY 161
5.4.3 TEACHING AND EDUCATION IN EAST- AND SOUTH-EAST ASIAN COUNTRIES 162
5.4.4 SUMMARY 165
5.4.5 TEACHING AND EDUCATION IN THE NORDIC COUNTRIES 165
5.4.6 SUMMARY 167
5.5 COMPARATIVE SUMMARY 167

CHAPTER 6 METHODOLOGY 170

6.1 MEASUREMENT INVARIANCE FOR TSE 170
6.2 MEASUREMENT INVARIANCE FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING 172
6.3 MEASUREMENT INVARIANCE FOR CLASSROOM CLIMATE 173
6.4 MEASUREMENT INVARIANCE FOR INSTRUCTIONAL LEADERSHIP 174
6.5 ANALYTICAL STRATEGY 175

CHAPTER 7 RESULTS RESEARCH QUESTION 2: INTERCULTURAL DIFFERENCES BETWEEN PREDICTORS OF TEACHER'S SELF-EFFICACY 176

7.1 RESEARCH QUESTION AND HYPOTHESES 177
7.1.1 TO WHAT EXTENT IS TSE IN INSTRUCTION PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN EACH OF THE THREE COUNTRY CLUSTERS? 177
7.1.1.1 Anglo-Saxon countries 178
7.1.1.2 Nordic countries 180
7.1.1.3 East- and South-East Asian countries 181

7.1.2 SIMILARITIES AND DIFFERENCES IN THE COUNTRY CLUSTERS FOR TEACHERS' SELF-EFFICACY IN INSTRUCTION 182
7.1.2.1 Teacher characteristics 182
7.1.2.2 Classroom characteristics 185
7.1.2.3 Principal characteristics 188
7.1.2.4 School characteristics 189
7.1.3 TO WHAT EXTENT IS TSE IN STUDENT ENGAGEMENT PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN THE THREE COUNTRY CLUSTERS? 191
7.1.3.1 Anglo-Saxon countries 193
7.1.3.2 Nordic countries 194
7.1.3.3 East- and South-East Asian countries 194
7.1.4 SIMILARITIES AND DIFFERENCES IN THE COUNTRY CLUSTERS FOR TEACHER'S SELF-EFFICACY IN STUDENT ENGAGEMENT (ETSE) 196
7.1.4.1 Teacher characteristics 196
7.1.4.2 Classroom characteristics 197
7.1.4.3 Principal characteristics 200
7.1.4.4 School characteristics 201
7.1.5 TO WHAT EXTENT IS TSE IN CLASSROOM MANAGEMENT PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN THE THREE COUNTRY CLUSTERS? 202
7.1.5.1 Anglo-Saxon countries 202
7.1.5.2 Nordic Countries 205
7.1.5.3 East- and South-East Asian countries 205
7.1.6 SIMILARITIES AND DIFFERENCES IN THE COUNTRY CLUSTERS FOR TEACHER'S SELF-EFFICACY IN CLASSROOM MANAGEMENT (MTSE) 207
7.1.6.1 Teacher characteristics 207
7.1.6.2 Classroom characteristics 209
7.1.6.3 Principal characteristics 211
7.1.6.4 School characteristics 212
7.2 SUMMARY 213
7.2.1 DIFFERENCES IN THE EXPLANATION OF VARIATION ACROSS LEVELS 214
7.2.2 DIFFERENCES IN THE ASSOCIATION OF THE PREDICTORS 214

CHAPTER 8 DISCUSSION AND CONCLUSIONS 216

8.1 MAIN RESULTS AND CONTRIBUTIONS TO THE FIELD OF RESEARCH ON TEACHERS' SELF-EFFICACY 216
8.1.1 RESEARCH QUESTION ONE: TEACHER'S SELF-EFFICACY IN AN INTERNATIONAL CONTEXT 217
8.1.2 RESEARCH QUESTION TWO: COUNTRY CLUSTER COMPARISONS OF PREDICTORS OF THREE DOMAINS OF TEACHER'S SELF-EFFICACY 219
8.2 STRENGTHS AND LIMITATIONS OF THE STUDY 220
8.2.1 A COMPREHENSIVE SURVEY DESIGN 221
8.2.2 MISSING DATA TREATMENT 223
8.2.3 A MULTI-FACETED MEASURE 224
8.2.4 A STATE-OF-THE-ART INSTRUMENT 224
8.2.5 A MULTILEVEL PERSPECTIVE 225
8.2.6 BRINGING TOGETHER TEACHERS' AND PRINCIPAL'S PERSPECTIVES 226
8.2.7 BROADENING THE RESEARCH FOCUS 227
8.3 DIRECTIONS FOR FUTURE RESEARCH 228
8.3.1 EXPANSION INTO OTHER COUNTRIES 228
8.3.2 USING LINKS WITH OTHER CLASSROOM OBSERVATION TOOLS 229
8.3.3 REACHING OUT TO NEW METHODS 230
8.3.4 COLLECTING INDEPENDENT DATA 231
8.4 IMPLICATIONS FOR EDUCATION POLICY 232
8.4.1 POSSIBLE IMPLICATIONS FOR TEACHER TRAINING 232
8.4.2 SUGGESTIONS FOR EDUCATION POLICY 234
8.5 SUMMARY 236

REFERENCES 237

APPENDICES 253

List of Tables

TABLE 1 DESCRIPTIVE STATISTICS FOR TEACHERS, SCHOOLS AND COUNTRIES 61

TABLE 2 MODEL FIT FOR ALL THREE SUB SCALES OF TSE 77

TABLE 3 FACTOR LOADINGS FOR CFA OF ALL COUNTRIES ON A THREE-FACTOR TSE SCALE (OECD, 2014) SINGLE-LEVEL 77

TABLE 4 TSE: TWO-LEVEL CFA (TEACHERS NESTED IN SCHOOLS) CARRIED OUT IN EACH COUNTRY SEPARATELY (MODEL FIT INDICES FOR EACH COUNTRY) 78

TABLE 5 MODEL FIT FOR MEASUREMENT INVARIANCE OF TSE ACROSS 32 COUNTRIES (C.F. OECD, 2013) 79

TABLE 6 RELIABILITY STATISTICS FOR EACH SUB SCALE OF TSE BY COUNTRY 80

TABLE 7 CFA FOR INSTRUCTIONAL STYLE OF LEADERSHIP BY COUNTRY (C.F. OECD 2013) 82

TABLE 8 RELIABILITY STATISTICS FOR INSTRUCTIONAL STYLE OF LEADERSHIP (C.F. OECD, 2013) 83

TABLE 9 MEASUREMENT INVARIANCE RESULTS FOR INSTRUCTIONAL LEADERSHIP (C.F. OECD, 2013) 84

TABLE 10 CFA FOR CLASSROOM CLIMATE BY COUNTRY (C.F. OECD, 2013) 85

TABLE 11 RELIABILITY AND DESCRIPTIVE STATISTICS FOR CLASSROOM CLIMATE 86

TABLE 12 MEASUREMENT INVARIANCE FOR CLASSROOM CLIMATE ACROSS 32 COUNTRIES (C.F. OECD 2013) 87

TABLE 13 MODEL FIT FOR CFA BY COUNTRY FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING (C.F. OECD 2013) 87

TABLE 14 MEASUREMENT INVARIANCE RESULTS FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING ACROSS 32 COUNTRIES (C.F. OECD, 2013) 88

TABLE 15 RELIABILITY AND DESCRIPTIVE STATISTICS FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING 89

TABLE 16 RE FOR DIFFERENT MS AND ΛS (OWN TABLE AFTER HORTON & LIPSITZ, 2001 P. 246) 103

TABLE 17 MISSING VALUES PER VARIABLE 103

TABLE 18 RESULTS FOR TSE IN STUDENT ENGAGEMENT 116

TABLE 19 RESULTS FOR TSE IN INSTRUCTION 117

TABLE 20 RESULTS FOR TSE IN CLASSROOM MANAGEMENT 118

TABLE 21 MEANS OF THE COUNTRY VALUES BY CLUSTER AND COUNTRY 154

TABLE 22 MEASUREMENT INVARIANCE FOR CONFIGURAL AND METRIC INVARIANCE FOR ETSE, ITSE AND MTSE FOR ALL THREE COUNTRY CLUSTERS 171

TABLE 23 MI FOR CONFIGURAL AND METRIC INVARIANCE FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING 172

TABLE 24 2 MI FOR CONFIGURAL AND METRIC INVARIANCE FOR DISCIPLINARY CLASSROOM CLIMATE 173

TABLE 25 MI FOR CONFIGURAL AND METRIC INVARIANCE FOR INSTRUCTIONAL LEADERSHIP STYLE 174

TABLE 26 NUMBER OF TEACHERS AND SCHOOLS IN EACH COUNTRY CLUSTER 176

TABLE 27 RESULTS FOR TSE IN INSTRUCTION FOR ALL THREE COUNTRY CLUSTERS 179

TABLE 28 RESULTS FOR TSE IN STUDENT ENGAGEMENT FOR ALL THREE COUNTRY CLUSTERS 192

TABLE 29 RESULTS FOR TSE IN CLASSROOM MANAGEMENT FOR ALL THREE COUNTRY CLUSTERS 204

TABLE 30 DESCRIPTIVE STATISTICS FOR INDEPENDENT VARIABLES ACROSS 32 OECD COUNTRIES 254

TABLE 31 DESCRIPTIVE STATISTICS FOR PREDICTORS IN RQ2; ANGLO-SAXON COUNTRIES 255

TABLE 32 DESCRIPTIVE STATISTICS FOR PREDICTORS IN RQ2; NORDIC COUNTRIES 256

TABLE 33 DESCRIPTIVE STATISTICS FOR PREDICTORS IN RQ2; ESEA COUNTIRES 257

TABLE 34 CRONBACH'S ALPHA FOR EACH LATENT VARIABLE PER COUNTRY CLUSTER 258
TABLE 35 CORRELATION MATRIX FOR PREDICTOR VARIABLES OF RQ1 AND RQ2 259
TABLE 36 DIFFERENTIAL RESULTS FOR ITSE 260
TABLE 37 DIFFERENTIAL RESULTS FOR ETSE 260
TABLE 38 DIFFERENTIAL RESULTS FOR MTSE 260

List of Figures

FIGURE 3-1 BUILD-UP OF THE MLM FOR THE FIRST RESEARCH QUESTION 68
FIGURE 3-2MULTILEVEL STRUCTURE OF TEACHER SELF-EFFICACY 73
FIGURE 3-3 MISSINGNESS PER VARIABLES, CASES AND VALUES 106
FIGURE 3-4 MISSING VALUE PATTERNS 107
FIGURE 3-5 10 MOST FREQUENTLY OCCURRING PATTERNS 108
FIGURE 3-6 MISSINGNESS PER VARIABLES, CASES AND VALUES 108
FIGURE 3-7 MISSING VALUE PATTERNS 109
FIGURE 3-8 10 MOST FREQUENTLY OCCURRING PATTERNS 110
FIGURE 3-9 MISSINGNESS PER VARIABLES, CASES AND VALUES 110
FIGURE 3-10 MISSING VALUE PATTERNS 111
FIGURE 3-11 10 MOST FREQUENTLY OCCURRING PATTERNS 112
FIGURE 3-12 MISSINGNESS PER VARIABLES, CASES AND VALUES 112
FIGURE 3-13 MISSING VALUE PATTERNS 113
FIGURE 3-14 10 MOST FREQUENTLY OCCURRING PATTERNS 114

List of Equations

EQUATION 1 GRAND-MEAN CENTERING 63

EQUATION 2 INTRA-CLASS CORRELATION ON THE TEACHER LEVEL 64

EQUATION 3 INTRA-CLASS CORRELATION ON THE SCHOOL LEVEL 64

EQUATION 4 INTRA-CLASS CORRELATION COUNTRY-LEVEL 65

EQUATION 5 MODEL 1 RQ1: TEACHER PREDICTORS 66

EQUATION 6 MODEL 2 RQ1: CLASSROOM PREDICTORS 66

EQUATION 7 MODEL 3 RQ1: PRINCIPAL PREDICTORS 66

EQUATION 8 MODEL 4 RQ1: SCHOOL PREDICTORS 67

EQUATION 9 DATA MISSING COMPLETELY AT RANDOM 92

EQUATION 10 DATA MISSING AT RANDOM 93

EQUATION 11 DATA MISSING NOT AT RANDOM 94

EQUATION 12 RELATIVE EFFICIENCY OF IMPUTED VALUES 103

EQUATION 13 FULL MODEL RQ2: TEACHER-, CLASSROOM-, PRINCIPAL- AND SCHOOL PREDICTORS 175

List of Appendices

Appendix 1: CUREC letter of approval	253
Appendix 2: Descriptive Statistics RQ1	254
Appendix 3: Descriptive Statistics RQ2	255
Appendix 4: Correlation Matrix of Predictor Variables	259
Appendix 5: QQ-Plots for teacher level variables	260
Appendix 6: Significance tests for differences in the predictors of RQ2	265

Abbreviations

App.	Appendix
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
DV	Dependent Variable
ECEC	Early Childhood Education and Care
Fig.	Figure
FIML	Full Information Maximum Likelihood
GDP	Gross Domestic Product
GLOBE	Global Leadership study
ICC	Intra Class Correlation
INES	Indicators of Education Systems
ISCED	International Standard Classification of Education
IMHE	Institutional Management in Higher Education
IV	Independent Variable
MAR	Missing At Random
MCAR	Missing Completely At Random
MNAR	Missing Not At Random
MI	Multiple Imputation
ML	Maximum Likelihood
NPM	National Project Manager
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
RE	Relative Effectiveness
RMSEA	Root Mean Square Error of Approximation
RQ	Research Question
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences
SRMR	Square Root Mean Error
Tab.	Table
TALIS	Teaching and Learning International Survey
TLI	Tucker Lewis Index
TSE	Teacher self-efficacy
MTSE	Teacher self-efficacy in Classroom Management
ITSE	Teacher self-efficacy in Instruction
ETSE	Teacher self-efficacy in Student Engagement
VET	Vocational Educational Training

DISSERTATION OVERVIEW

The following chapter aims to provide a brief overview of the dissertation by summarising each chapter in a short paragraph.

Chapter 1 contains the introduction to the thesis and provides the rationale for the research as well as the scope of the research. It highlights the importance of the concept of teacher's self-efficacy with respect to previous research findings and how it relates to education theory and praxis. Moreover, it introduces the scope of the research by mapping the different data sources and how they can be used to advance the field of research on teacher's self-efficacy.

Chapter 2 provides the theoretical framework by explaining the theories that led to forming the concept of teacher's self-efficacy and how they relate to the concept. Theories mentioned include the *Social learning theory, Social cognitive theory* and *Attribution theory*. Attention is being paid to the various ways in which teachers' self-efficacy has been measured in the past and the current tripartite structure embracing teacher's self-efficacy in student engagement, instruction and classroom management. It furthermore contains sections on past investigations of teacher's self-efficacy with respect to teacher-, classroom-, principal- and school characteristics and thus provides the rationale for the choice of the predictors used in the statistical models later in the thesis.

Chapter 3 contains a description of the methodology used to prepare the data for analysis and the statistical analysis itself. It mentions *ethical considerations,* explains *secondary quantitative data analysis*, the *TALIS data*, its properties and set-up of the questionnaires and a reflection on the use of OECD data in an education context. Furthermore, it explains the analytic procedure, description and quality assessment of the variables used, the multilevel structure of the data and analytical requirements it has. Substantive attention is being paid to the analysis of missing data patterns in the data sets as well as appropriate treatment of missing values and their imputation in MPlus.

Chapter 4 is the first of two chapters containing the results and interpretations of the two big research questions, around which this thesis is built. It is structured according to the three dimensions of teacher's self-efficacy that have been investigated previously: To which extent teacher's self-efficacy in student engagement, instruction and classroom management is predicted by teacher-, classroom-, principal- and school characteristics in the 32 OECD and partner countries in the TALIS data-set. It furthermore contains sections, in which the results of the three domains of teacher's self-efficacy are compared.

Chapter 5 provides the theoretical framework for research question two. Although the literature on teacher's self-efficacy for research question one also applies here, there needs to be a theoretical framework for choosing the three country clusters containing Anglo-Saxon, Nordic and East- and South-East Asian countries. Thus, this chapter contains the rationale for choosing the clusters according to shared characteristics of their education systems, country values and to which extent they prevail in the countries as well as a section on how teaching styles and schooling are related to cultural settings.

Chapter 6 is a complementary methodology chapter that contains additional analyses that became necessary for the second big research question. The main methodology described in Chapter 5 also applies to research question two, but due to the selection of sub-samples from the main data-set some additional analyses regarding the quality of the scales were conducted.

Chapter 7 contains the results and interpretations for research question two that investigated cultural differences between predictors of teacher's self-efficacy within the Anglo-Saxon, Nordic and East- and South-East Asian country clusters across the three domains of teacher's self-efficacy in student engagement, instruction and classroom management. The interpretations were drawn from the literature on teacher's self-efficacy and the cultural framework established beforehand.

Chapter 8 concludes the thesis by bringing together the first and second research question in the discussion and conclusion chapter. It places the current research in the landscape of

research on teacher's self-efficacy by highlighting its contributions to the field. Furthermore, it discusses the strengths and limitations of the study by pointing towards the challenges along the way and describing how they were dealt with. Then it gives directions for future research that emerge from the thesis and draws implications for education policy and teacher training that the results suggest.

CHAPTER 1 INTRODUCTION

This chapter introduces the thesis by giving an overview of key findings in research on teacher's self-efficacy that constitute the field and led to the research questions. It also describes the data-set used and gives an outlook on the expectations and limitations the data brings to the study. Chapter 2 provides a more detailed account of the theoretical framework and discussion of the literature.

1.1 RATIONALE FOR THE RESEARCH

The environment of teaching, and thus the role of the teacher as a facilitator of such an environment, has been paid considerably more attention in research over the past decades (Klassen, Tze, Betts, & Gordon, 2010). One factor that is assumed to have an influence in creating a stimulating and productive learning environment as part of the teacher's characteristics is the teacher's self-efficacy (TSE) (Tschannen-Moran & Hoy, 2001). The interest in TSE goes back to Albert Bandura, who first introduced the concept of self-efficacy as a belief in *"one's capability to accomplish a given level of performance"* (Bandura, 1986, p. 391). It describes a person's conviction of being equipped with the necessary knowledge, abilities and training to successfully perform a certain task. With respect to teachers, it has been further specified as *"teachers' belief or conviction that they can influence how well students learn, even those who may be difficult or unmotivated"* (Guskey & Passaro, 1994, p. 4). So, TSE focuses on the teacher's own ideas of how successful their teaching will be, regardless of possibly difficult circumstances. Previous research has shown that characteristics of teachers with a high sense of self-efficacy are, for example, to expect success in difficult situations in the classroom or to motivate students (e.g. Giallo & Little, 2003; Woolfolk, Rosoff, & Hoy, 1990), to be less concerned with strictly controlling pupils (Woolfolk & Hoy, 1990) or to show openness toward new methods (Guskey, 1988). Furthermore, teachers who feel they are self-efficacious are more likely to report they are supported by an authoritative and understanding principal who sets clear goals (Hoy & Woolfolk, 1993; Walker & Slear, 2011) or teach in private schools (Chester & Beaudin, 1996). However, although the concept is nowadays highly recognised internationally, most research so far has been carried out in the North American context (Klassen et al., 2010).

This research seeks to contribute to the study of TSE in an international context. The data being used in this study stems from the Organisation for Economic Cooperation and Development (OECD), an organisation that seeks to enhance the quality of life and economies of its member countries around the world (OECD, 2013a). It has identified education as one of the major contributing factors in order to respond to the challenges and difficulties modern societies are facing: "[the] *financial crisis, meeting development goals, encouraging green growth and responding to climate change, ageing societies and the knowledge economy"* (OECD, 2011). Thus, it carries out various empirical studies in order to provide a system of *" indicators [as] a strong evidence base for international comparisons of all aspects of education systems"* (ibid. p. 1) that provides insight into the state of the educational systems of its member countries, including; Pre-school and school (Starting Strong, PISA, TALIS, Sustainable Schools and OECD Review on Policies to Improve the Effectiveness of Resource Use in Schools), vocational education (VET, OECD Skills Strategy, OECD Skills Outlook, Innovation Strategy for Education and Training and Education and Social Progress), higher education (OECD Higher Education Programme (IMHE) and Assessment of Higher Education Learning Outcomes (AHELO) as well as adult learning (Survey of Adult Skills (PIAAC)) (OECD, 2011). Accordingly, these studies aim to provide educational policy makers with a rich source of data for evidence-based decision-making in order to implement changes based on a solid empiric ground (ibid.).

However, with respect to the educational context, in which this study is placed, it should be mentioned that the OECD views education rather as a means for economic ends than a value in itself, as it says: *The mission of the Organisation for Economic Co-operation and Development is to promote policies that will improve the economic and social well-being of people around the world* (OECD, 2013a). Therefore, it provides an economic-focused view on education, which influences the choice of topics it surveys. It is important to bear this in mind for the further considerations of this study.

The environment in which teaching and learning take place is investigated by TALIS (Teaching and Learning International Survey) which was designed to "*increase the amount of international information available to OECD countries on teachers, teaching and the impact*

that teachers can have on student learning" (OECD, 2013c, p. 29). Together with the participating countries, the OECD has agreed on the following six focus areas:

1. School leadership
2. Teacher training and in-service professional development/initial teacher education
3. Teacher appraisal and feedback
4. School climate and ethos
5. Teachers' pedagogical beliefs
6. Teachers' pedagogical practices

Within these areas, teachers' self-efficacy has been identified as an important concept that has inter alia (i.a. in the following) been the focus of the survey. With respect to the above-mentioned importance of the concept, the availability of the TALIS data set will enable me to investigate teachers' self-efficacy on an international comparative level and thus to address a gap in the literature as well as to contribute to the understanding of the way TSE and the influences that shape it may vary between countries. This will be done by first, investigating predictors of teacher's self-efficacy across 32 OECD countries. Second, by comparing those predictors across three country clusters comprising Anglo-Saxon, Nordic and East- and South-East Asian countries, cultural differences and their relation to teacher's self-efficacy will be investigated.

1.2 SCOPE OF THE RESEARCH

TALIS 2013 has been carried out in 23 OECD countries and 8 partner countries by investigating the teachers of 15 year old students and their principals (OECD, 2013c). Accordingly, it compiles the information on about 104,000 teachers and 6,400 principals. However, although the nested structure of the data (teacher nested in schools nested in countries) and the choice of scales, which are mainly continuous, provides the opportunity to address important research questions about TSE by applying complex statistical methods and a wide

range of analytical models, the data only provide information for the teachers and classrooms as well as principals and schools. Thus, to gain insight into cultural differences of the teachers' workplace internationally, additional information to form an interpretive framework will be drawn from intercultural research done by House et al. (House, Hanges, Javidan, Dorman, & Gupta, 2004) in the GLOBE study that mapped cultural beliefs and values in 64 countries or from Hofstede (1986), whose research focuses on leadership values on a comparative international basis. Furthermore, a sample of Anglo-Saxon, Nordic and East- and South-East Asian, which were grouped according to similarities of their education systems (Bulle, 2011), will be compared to gain understanding in the different work environments of the teachers in which TSE develops. With this it is hoped to broaden the understanding of the environment in which teachers work in relation to their cultural beliefs and characteristics of the school systems.

With respect to these possibilities, this study aims to contribute to research in teacher self-efficacy in the following ways:

1. As TSE is seen as an important influencing factor on the quality of teaching and thus indirectly contributes to the economic success and wellbeing of people, the study aims to enhance understanding of the factors that show differences in different school systems that operate in different societies and cultures. Therefore, between and within country analysis will be carried out to see where differences occur and try to explain them.

2. In this study, TSE as a psychological construct will be analysed in conjunction with cultural, sociological and organisational perspectives. Thus, it contributes to a holistic understanding of the broader context in which TSE develops, as it investigates several sources of influence that shape the teachers' work place (country, school, classroom as well as teachers' own characteristics).

3. The study will endeavour to provide new evidence about the influences that shape TSE in order to inform educational policy and evidence-based decision-making. Therefore, I will investigate the following two main themes: (1) How the TSE measures are predicted internationally, (2) How TSE is differently predicted across

country clusters that represent different education systems, namely the Anglo-Saxon, the Nordic and the East- and South East Asian.

CHAPTER 2 THEORETICAL FRAMEWORK AND LITERATURE REVIEW

This chapter will discuss the theoretical underpinnings of TSE and provide a review of the relevant literature addressing research conducted at the teacher level, classroom level, principal level and country level. Thus, it aims at establishing the theoretical framework of the research field and the multilevel context in which the research takes place. It will furthermore identify gaps in the literature that the study can address.

2.1 THEORIES LEADING TO TSE

The concept of teacher self-efficacy has its historic roots in three other social science theories: Social learning theory by Albert Bandura and Robert Sears (1971); Social cognitive theory by Albert Bandura (1977); and Attribution theory by Fritz Heider (1958) and Julian Rotter (1966). The following sections will introduce and discuss these theories with respect to their role in the development of teacher self-efficacy.

2.1.1 SOCIAL LEARNING THEORY

The first component of the concept of self-efficacy is Bandura's social learning theory. The belief underlying the theory is that human actions are influenced by both observing a role model, and the reaction of the environment towards the role model's behaviour (Ashford, LeCroy, & Lortie, 2009). In his 'Bobo doll experiment', Bandura exposed children to a scene in which a woman hit a Bobo doll and received two different reactions: one group of children saw her being rewarded for hitting the doll, the other group of children saw her being punished. The children were then left alone in the room with the Bobo doll and their behaviour was observed: those children who watched the woman receiving praise for hitting the doll were more likely to hit the doll whereas those children who watched the woman receiving punishment, did not hit the doll as often. Those children who successfully adopted the role model's behaviour fulfilled four pre-conditions (Bandura, 1971):

1. They paid *attention*.
2. They used *retention* as the active memory of the observed.

3. They *reproduced* the observed behaviour by using their own mental and physical abilities.

4. They experienced *motivation* as an incentive to expect a certain outcome by imitating the observed behaviour.

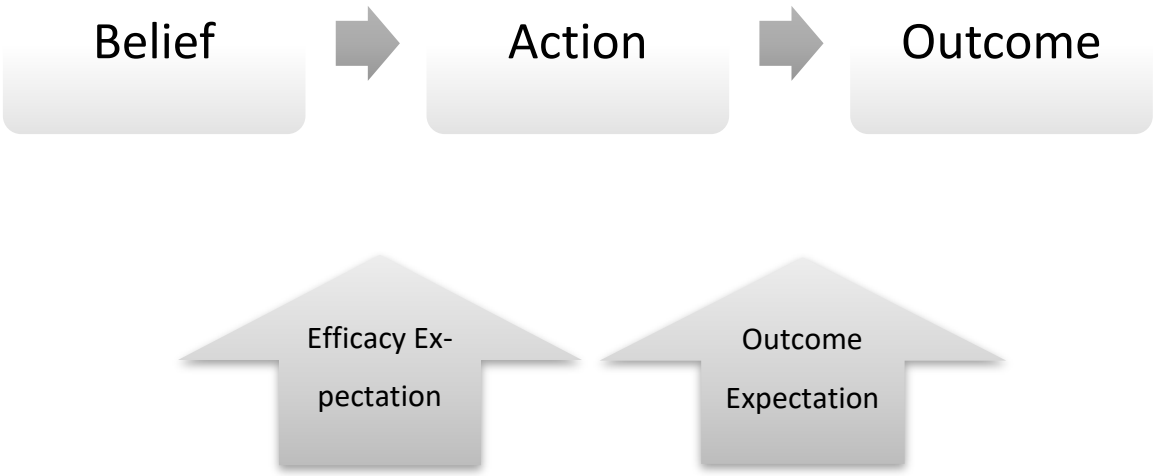
For the concept of self-efficacy, Bandura's two conclusions that (1) behaviour is influenced by role models and (2) humans adopt their behaviour with respect to the reactions they expect from their environment are fundamental to understanding how and why people act as they do in certain situations. Accordingly, when applied to TSE, teachers who are exposed to a successful, or by contrast to a failing role model that they strongly identify with (e.g. principal or more experienced colleague) may draw conclusions that lead either to an increase or to a decrease of their own sense of efficacy (Bandura, 1997).

2.1.2 SOCIAL COGNITIVE THEORY

The second theory by Bandura, which has influenced the concept of self-efficacy, is social cognitive theory. Social cognitive theory is a further development of the social learning theory mentioned above that takes into account the process a person is going through once a belief is turned into an action that leads to a desired outcome (Leonard, 2002).

Accordingly, Bandura explains the efficacy expectations of a person in contrast to the outcome expectations of the person as a two-step-process: First, efficacy beliefs account for the degree to which people believe they can influence an action in order to achieve a certain outcome. Second, outcome expectations account for the degree to which people believe their action can influence this outcome. Thus, the efficacy belief stands in a logical order before the outcome belief (Bandura, 1977). Figure 1 presents an illustration of this sequencing.

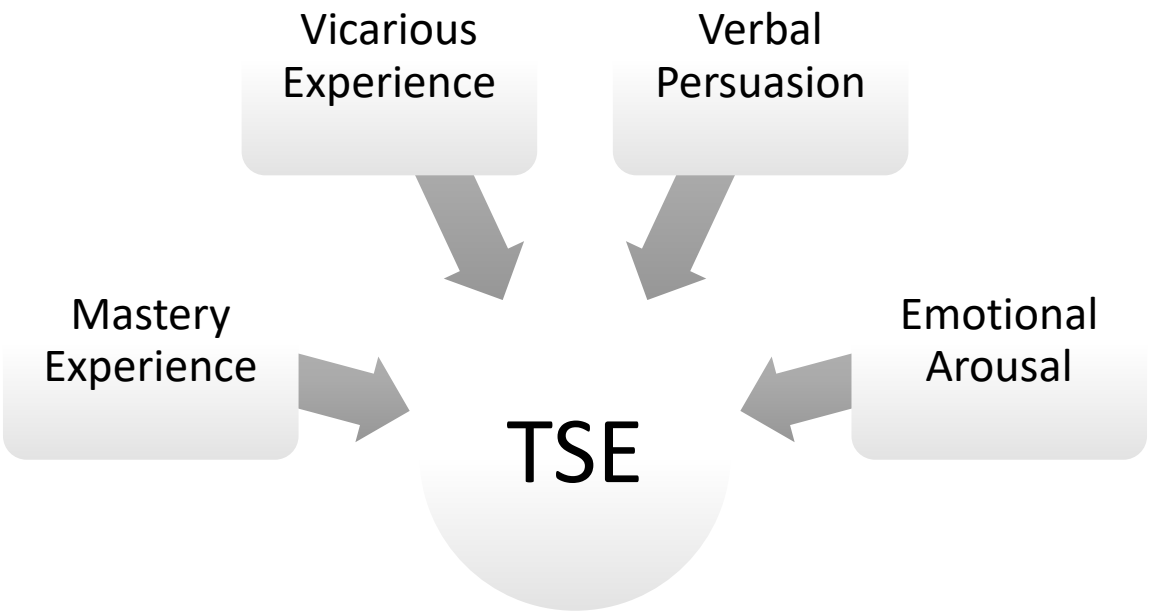
Fig. 1: The relationship between Efficacy- and Outcome Expectations (own graphic)



As these efficacy beliefs are emerging in a particular environment, they are not necessarily stable, but can vary across three different dimensions (ibid.): Magnitude, generality and strength. *Magnitude* describes, on which level of difficulty the task operates, *generality,* if the task requires rather general or mastery knowledge and *strength,* the degree, to which a person believes the achieved action will be successful. Key to this is the person's self-reflective capability that evaluates the chances of future success with respect to past experiences. But thereby a causal order cannot be defined, because mastery expectations and outcome beliefs are thought to influence each other reciprocally. In addition, efficacy beliefs are held to be influenced by the following four sources: Performance accomplishment, vicarious experience, verbal persuasion and emotional arousal (see Fig. 2). *Performance accomplishments* or *mastery experiences* are held to be generated by frequent successful actions that have a positive effect on efficacy belief, whereas failures have a negative impact on the efficacy belief. But once a particular efficacy belief is established in a certain situation, it can be generalised to other situations. From all four sources of self-efficacy, mastery experience has been shown being a powerful source of self-efficacy in previous research, whereas the following three still need further investigation (Henson, 2002). *Vicarious experience*, or learning through a role model, is the second factor hypothesised to enhance efficacy beliefs: If the role model succeeds in conducting a presumably difficult task, self-efficacy beliefs are expected to rise. The more someone identifies with the role model, the larger is the learning effect and the more someone identifies with the

role model. The third of the four sources, *verbal persuasion,* enhances self-efficacy beliefs by using positive verbal reinforcement as a form of feedback. This source builds on a person's expectation of the outcome of a future event, as previous positive feedback establishes a positive attitude towards the result. The last source, *emotional arousal,* builds on the power that emotions have on the expectations of success or failure: If positive feelings are associated with a situation, the person is more likely to perceive high self-efficacy, whereas negative emotions can lead to stress and thus might enhance the possibility to fail as a result of lower expectancies (ibid. & Merrell, 2003). Accordingly, the development of self-efficacy can be seen as a product of dynamics between the personal, social, situational and temporal conditions in which a person acts (Jarvis, 2003). Figure 2 presents an illustration of the various sources of influence on TSE identified in Social cognitive theory.

Fig. 2 Sources of Teacher Self-Efficacy beliefs (own graphic)



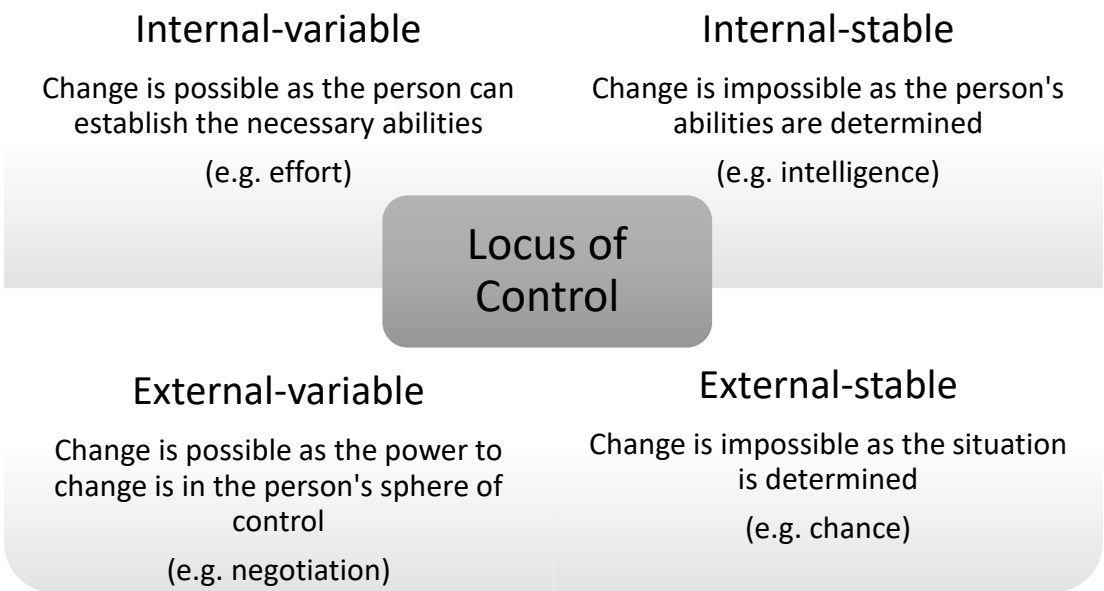
2.1.3 ATTRIBUTION THEORY

The third theory, on which the concept of self-efficacy is built, is Heider's (1958) attribution theory, which was later expanded by Rotter (1966). Heider assumes that human beings attribute success or failure either to dispositional or situational sources in order to logically justify their behaviour. Dispositional sources are explained by a person's characteristics

and can include qualities such as effort, motivation or other internal abilities (Augoustinos, Walker, & Donaghue, 2006). Here, the power to change or influence a situation is proposed to lie within the person's sphere of control. Situational sources, in contrast, originate from a person's environment and can be represented by chance, institutions or powerful people (ibid.). Accordingly, the power to change or influence a situation lies outside the person's sphere of control. It has been shown that people tend to overestimate the personal, but underestimate the contextual impact of situations, which is called the 'fundamental attribution error' (Ross, 1977).

Later, the concept was further advanced by Rotter, who added a second dimension: The first dimension, which he calls '*locus of control*', is identical with Heider's idea of attribution. The newly added second dimension is *stability*: The internal or external source of success/failure can either be stable or variable. This leads to four loci of control: internal-variable (the source of failure/success lies within the person and can be changed through his or her own effort), internal-stable (the source of failure/success lies within the person, but cannot be influenced as the person's abilities are determined) external-variable (the source of failure/success lies within the situation and can be influenced by the person's effort) and external-stable (the source of failure/success lies within the situation but cannot be influenced or changed). In Figure 3 illustrates these four positions on Locus of Control.

Fig. 3: Loci of control (own figure)



2.1.4 SUMMARY

From these three authors, two different approaches to measure self-efficacy originate. Bandura's social learning theory and social cognitive theory contribute the two dimensions of efficacy- and outcome expectations: Efficacy beliefs are *''beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments''* (Bandura, 1997, p. 3). Thus, these beliefs inform the positive expectancy to be capable of future events. Outcome efficacy, in contrast, informs the belief to be able to perform a task at a given level (Bandura, 1986) and thus to bring forth a certain outcome. Furthermore, Bandura formulates four sources of self-efficacy (performance accomplishment, vicarious experience, verbal persuasion and emotional arousal) that will provide the theoretical model for the sources of the teacher's self-efficacy in the present study. This has been picked up by Gibson & Dembo (1984), who confirmed Bandura's distinction of a person's efficacy versus outcome expectation: Personal teaching efficacy follows the assumption that the teacher is able to perform the necessary tasks to stimulate change, whereas general teaching efficacy reflects the extent to which a teacher believes students can be taught regardless of other external factors e.g. their SES or school conditions.

With respect to teachers, Rotter's 'locus of control' can be interpreted as the teachers' belief either that the source of their pupils' educational failure or success lies within their own sphere of control (= internal locus) and thus they can influence their pupils' educational success significantly. Or, the possibility to influence educational success lies outside their sphere of control, e.g. in the parental home, the pupils' cognitive ability or school climate (= external locus). In this paradigm, the teachers' belief of their chances to bring about change in their pupils' learning behaviour through teaching is very limited. This has been empirically supported by Guskey & Passaro (1994) who found two scales of TSE representing what Rotter has described as an internal versus an external locus of control.

2.2 INDIVIDUAL CHARACTERISTICS OF TSE

Over approximately the past 40 years, a rich body of literature investigating what is called 'teacher's self-efficacy' has been developed. Building on the above-mentioned theories, a myriad of instruments has been developed that has each been focussing on other aspects

of the initial theories; often only for the purpose of one study. For instance, Morris et al. (2017) found 19 different instruments used in 67 studies for the suggested period.

Thus, the construct of TSE has widely been discussed in the literature and still is among researchers (Tschannen-Moran & Hoy, 2001). But still, there is no final consensus on how the concept should be applied in research, although there is convergence around the measurement suggested by Tschannen-Moran & Woolfolk Hoy (2001). Chesnut & Burley (2015) in their meta-analysis pointed out the importance of measuring the concept conceptually accurately so that associations can be measured properly. In order to show the variety of measurements that exist, exemplarily two constructs of each theoretical strand will briefly be introduced in the following.

2.2.1 EXAMPLE OF TWO MEASUREMENTS USING ROTTER'S 'LOCUS OF CONTROL'

The first measurements of TSE were informed by Rotter's 'locus of control'. So, measurements building on Rotter will be introduced first, followed by measurements building on Bandura or combining both.

2.2.1.1 The RAND-measurement

The initial measure of TSE consisted of two items, which were used by the RAND[1]-researchers Armor et al. (1976), who used Rotter's concept of the external and internal locus of control in order to (1) assess the strengths of the teacher's belief that the environment would influence the teaching success more than the teacher's own effort[2] (=external locus). This is the so-called 'general teaching efficacy' and explores to what extent teachers rate the general power they have in the classroom to bring about change in their students'

[1] For more information about the RAND-corporation see http://www.rand.org/about.html

[2] Rand item 1. ''When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment.''

Rand item 2. ''If I really try hard, I can get through to even the most difficult or unmotivated students."

lives by teaching. Furthermore, it (2) assessed whether teachers belief their own course of action could overcome given external circumstances like SES or motivation (= internal locus). The notion, that teachers believe they possess the ability to influence student learning in a positive way, is also called 'personal teaching efficacy' (PSE).

2.2.1.2 Guskey's concept of 'Responsibility of student achievement'

One scale, building on Rotter's loci of control, was developed by Guskey (1981): The concept of 'Responsibility for student achievement' (RSA). Teachers were given a list of 30 different situations, for which two different types of solutions were provided: One showed an internal, one an external locus of control. Furthermore, 15 items indicated a negative, 15 a positive responsibility for student achievement. The teachers were asked to distribute a total of 100% between them, according to how much they agreed to each of the two solutions. From a factor analysis, two factors emerged: One showing responsibility for student success (R+), one for student failure (R-). These two scales were positively correlated with the initial RAND-items, but only weakly (r = 0.20) with each other (Tschannen-Moran & Hoy, 2001). Aside from the initial study, the construct was used by Guskey (1987), in order to investigate contextual variables that might affect TSE and in order to investigate the relationship between TSE and the implementation of new instructional methods (Guskey 1988). Although the scale was carefully developed and fitted into the research, it was not used by researchers other than Guskey himself (Tschannen-Moran & Hoy, 2001).

2.2.1.3 Rose & Medway's contextualisation of TSE

Rose & Medway (1981) tried to allocate the teacher locus of control by letting them choose between two possible explanations for a classroom situation. The list consisted of 28 items; 14 indicating failure, 14 indicating success in the classroom. One option located the responsibility within the teacher, the other one within the situation. The authors argue, that their construct is a better measurement for TSE than the original RAND-items, because it is more context specific. But nevertheless, further research showed that the construct was positively weakly (r = 0.11) to moderately (r = 0.41) correlated to the RAND-measures (Parkay, Greenwood, Olejnik, & Proller, 1988). The main difference to the scale developed by Guskey is that the respondents have to decide for one of the options instead of showing their

degree of agreement for both of them. Besides the initial study that validated the scale and found that TLC can predict classroom management and the teacher's willingness to adopt new teaching strategies, it was used by Rose and Medway (Rose & Medway, 1981a) again, who found that teachers with an internal locus of control teach higher achieving classes than teachers with an external locus of control.

2.2.2 EXAMPLES OF MEASUREMENTS USING BANDURA'S CONCEPT OF SELF-EFFICACY

The second conceptual strand is based on Bandura's concept of self-efficacy described in Chapter 2.1. In the following, two scales developed by Gibson and Dembo (1984), and Bandura himself will be introduced.

2.2.2.1 Gibson & Dembo's combination of the RAND-measure and Bandura's Outcome expectancies

Because researchers are still not sure how to adequately measure TSE, although the importance of self-efficacy on teaching has already been identified (Gibson & Dembo, 1984), Gibson and Dembo developed a scale of TSE that consists of 30 items in order to investigate this phenomenon more precisely. Their scale still builds on the two items of the RAND-measure, but includes Bandura's concept of outcome expectancies, as they argue that *"outcome expectancy would essentially reflect the degree to which teachers believed the environment could be controlled, that is, the extent to which students can be taught given such factors as family background, IQ and school conditions."* (ibid. p. 570). Thus, the gap in Rotter's locus of control, which only explains action-outcome beliefs, is extended by a measure of a person's belief in being able to conduct the necessary actions in order to bring about the desired change (Bandura, 1977; Gibson & Dembo, 1984). A factor analysis revealed two dimensions of TSE that fitted into the previous research: General teaching efficacy (GTE), which is the "*belief that any teacher's ability to bring about change is significantly limited by factors external to the teacher, such as the home environment, family background, and parental influences"* (ibid. p. 574) and personal teaching efficacy (PTE), which represents "*a teacher's belief that one has the skills and abilities to bring about student learning"* (ibid. p. 573). The GTE-factor appeared to correspond to Bandura's outcome-expectancy dimension, whereas the PTE-factor reflects his dimension of self-efficacy

(ibid.). Both factors were just moderately negatively correlated (r = -0.19), so both might represent related, but comparatively independent concepts. GTE turned out to explain 10.6% of the total variance, PTE 18.2%. The two factors that emerged from the factor analysis, indicate that TSE is a multi-dimensional construct that needs further theoretical development (ibid.). Furthermore, Tschannen-Moran & Hoy (2001) point out that although the factor structure was confirmed in several other studies, it did not appear to be stable, as several items loaded on both factors, when they were applied in different studies. Although this scale has been critiqued variously, it still is the most widely used as at least a starting point for research in TSE and has been adapted by several researchers to their specific purposes.

2.2.2.2 Bandura's conceptualisation of a measurement of TSE

The last scale explained here is an unpublished work by Bandura. It is included in the literature review although it has not yet been published, because it might be enlightening to reflect on how Bandura himself would apply his concept of self-efficacy to teachers in particular. Bandura (1997) argued that teachers' self-efficacy might not be stable across the various tasks the job entails. Thus, he developed a measurement with 30 items that comprises seven subscales which each represents a special task, which teachers have to perform in the classroom: efficacy to influence decision making, efficacy to influence school resources, instructional efficacy, disciplinary efficacy, efficacy to enlist parental involvement, efficacy to enlist community involvement, and efficacy to create a positive school climate. The items are measured on a 9-step Likert-scale with 5 anchor-items labelled 'nothing', 'very little', 'some influence', 'quite a bit', 'a great deal'. The items were phrased like questions, each beginning with 'How much…' (e.g. How much can you do to keep students on task on difficult assignments?). The content of many of the items is similar to other scales (e.g. Gibson & Dembo 1984, Woolfok & Hoy 2001), but the structure of the questionnaire as well as the phrasing of the items is different, as is the scale format.

Because it has not been used yet, there is no information available about the validity and reliability of the scale (Tschannen-Moran & Hoy, 2001). But nevertheless, it would be interesting to compare this scale to the ones used frequently in research for the following

reasons: (1) Previous research has shown, that different wording of the same conceptual ideas has a significant influence on the answering scheme of the respondents. (2) Tschannen-Moran & Hoy (2001) point out that none of the measurements developed so far, which range from rather complex to very simple ones, have met the balance between generality and specificity. It can be assumed that this questionnaire, as it combines frequently used measures with situation specific ones, might provide further insight.

2.2.2.3 Teacher self-efficacy scale by Schwarzer & Hallum, 2008

Another more recent conceptualisation of teachers' self-efficacy was proposed by Schwarzer (1995), who proposed a 10-item scale covering a variety of teaching tasks and classroom situations, which are namely job accomplishment, skill development on the job, social interaction with students, parents and colleagues and coping with job stress. Those items were measured on a 4-point Likert-scale ranging from 1 = not at all true to 4 = exactly true. There are three elements that characterise this instrument (Schwarzer & Hallum, 2008): 1., has an internal locus of control, as it attributes the causes of actions to the teachers; 2., it refers to future behaviours as it is based on teachers' estimates of how they will behave in a certain situation and 3., it is an operative construct, as it confronts the teachers with possible classroom situations and tasks and thus is a good predictor of actual behaviour.

In addition to using the initial two RAND-items described above, they created two items representing each of the four domains also described above, as i.e. 'I am confident in my ability to be responsive to my students' needs, even if I am having a bad day.' or 'I know that I can carry out innovative projects, even when I am opposed by sceptical colleagues.' The scale has been developed in accordance with Bandura's (1997) suggestion to keep an obstacle in each item (i.e. '… even if I am having a bad day.') to prevent overestimating one's self-efficacy by suggesting too easy tasks. Despite these strengths, this scale is only one-dimensional it is not possible to investigate TSE differentially for the variety of tasks the teachers have to perform in their daily work life, which is a limitation compared to other conceptualisations of TSE.

In the initial validation study, which had a longitudinal design measuring the self-efficacy of teachers in a number of German schools revealed a cronbach's alpha of α = .76 at timepoint 1 and α = .82 at timepoint 2. The retest-reliabilities for two and three years were r_{12} = .67, r_{23} = .76 and r_{34} = .78; respectively r_{13} = .65 and r_{14} = .61, which supports the stability of the scale (Schmitz & Schwarzer, 2002). However, it has only been used once in an English publication so far; all other applications of the scale happened in German research.

2.2.2.4 Teacher self-efficacy scale by Skaalvik & Skaalvik, 2017

As Schwarzer's scale is an example of a conceptualisation of teachers' self-efficacy that implies TSE as a general trait and thus is not able to measure different aspects of teaching separately, Skaalvik & Skaalvik (2007) proposed a scale that consists of six subscales. Those are: Instruction, Adapting Education to Individual Students' Needs, Motivating Students, Keeping Discipline, Cooperating with Colleagues and Parents, and Coping with Changes and Challenges and are measured on a 7-step Likert scale, which ranges from 1 = not certain at all to 7 = absolutely certain. The instrument was specifically developed for the Norwegian school context to capture the tasks Norwegian teachers in particular face. This scale, again, is built following Bandura's (1997) suggestions and is containing a difficulty in each item to prevent the teachers from overestimating their own self-efficacy by proposing too easy situations (i.e. `How certain are you that you can collaborate constructively with parents of students with behavioural problems?` instead of 'How certain are you that you can collaborate constructively with parents of students?'). In the studies used, the scale showed satisfactory Cronbach alpha-values of between .81-.90 for each sub-scale, and the subscales themselves showed correlations of between .29-.46 indicating there is no big overlap between them. This scale is the most comprehensive one used in research so far but has been specifically developed for the Norwegian education context. Thus, applying the instrument in other contexts would require situation specific changes to the scale.

2.2.2.5 The Ohio State Teacher Self-efficacy scale (OSTES)

The most up-to-date measure, which reaches comparatively most agreement among researchers, is the Ohio State Teacher Self-Efficacy scale developed by Tschannen-Moran & Hoy (2001). Building on mainly Bandura's items and taking on his suggestions on how to construct a TSE scale, as well as consulting other expert researchers, they have developed a scale for which factor analysis revealed three dimensions: efficacy for student engagement (8 items), efficacy for instructional strategies (7 items), and efficacy for classroom management (3 items); of which a version is used in the present study. It was the motivation to embrace the multifaceted tasks that teachers perform in the classroom and go beyond other, until then, established measures – to assess a broader range of teaching tasks. To establish construct validity, the OSTES was correlated with both Rand items ($r = .18$ and $r = .53$, $p < .01$), the personal teaching efficacy scale ($r = .64$, $p < .01$) general teaching efficacy scale ($r = .016$, $p < .01$) by Gibson & Dembo (1984). Those were chosen because of their rooting in theory and their use in previous research. The strongest correlations were with the personal teaching efficacy scale and the second Rand item, which is in line with the framing of assessing a teacher's own self-efficacy in relation to three frequently performed teaching tasks rather than investigating a teacher's belief about the abilities of teachers in general. Currently, this questionnaire provides the most established scale for research in TSE and seems to reach most consensuses among researchers (see Section 3.8.1. for the item wording and measurement properties of the construct). Apart from being grounded in theory and having shown validity, it is of acceptable length and thus shows usability. With respect to its international applicability, Klassen et al. (2009) conducted invariance tests across the three dimensions in five countries (Canada, North America, Cyprus, Korea and Singapore) across teachers from primary and secondary schools. They found that the scales show evidence of invariance in all countries and not only America, where the scale has been developed. Furthermore, the analysis revealed that the concept is applicable in several cultural settings, which tend to support different cultural practices, have different histories and their own school systems.

However, there are two major points of critique that have been uttered regarding the OSTES: First, against Bandura's (1997) recommendation, the OSTES does not contain any obstacles or difficult situations in their items so that teachers might overestimate their own sense of SE, as the task presented in the item might sound easier than it actually is. Second, although one advantage of the instrument is that it contains three scales and thus takes into account multiple tasks associated with teaching, it leaves out other relevant tasks of the teachers' daily work, as has been criticised by Skaalvik & Skaalvik (2017), who suggest six dimensions instead. Despite those relevant points of critique, it can be argued that the OSTES, which is used in the present study, is nonetheless suitable for the actual research purpose. Although only comprising three dimensions (SE in instruction, student engagement and classroom management) and leaving out aspects like collaboration or change in the school system, those dimensions reflect closely tasks associated with actual teaching and classroom situations. In contrast to the scale suggested by Schwarzer, which presents four different teaching tasks as part of one trait, it distinguishes between three aspects of teaching and treats them as separate phenomena. Thus, OSTES allows for a more in-depth investigation of TSE than Schwarzer but is more parsimonious than Skaalvik & Skaalvik. At the same time, it is of manageable length, which is an important feature in a large questionnaire like the one used in TALIS. Too long questionnaires can lead, for example, to respondent fatigue and result in drop-out, which causes problems, too. Hence, despite its shortcomings, the tripartite scale used by TALIS can be deemed suitable for investigating TSE in the classroom context, which is the aim of the present study.

2.2.3 SUMMARY

This overview just comprises four measures that differently build on the theoretical underpinnings of TSE. However, it might give a first hint at what some of the main challenges in research in TSE include. Although Tschannen-Moran & Hoy (2001) developed a widely accepted scale for TSE, the number and content domains of TSE are still discussed. Especially during the early years of research on TSE, a myriad of measurements has been used and only within the past 15 years, some kind of agreement between researches has been achieved on how to measure the concept. Thus, it should be borne in mind that what all

those different conceptualisation share, despite their differences, is being rooted in the seminal works of Rotter and Bandura, which serve as some kind of umbrella to unite the research done in this area.

2.3 PAST INVESTIGATIONS OF TEACHER'S SELF-EFFICACY

Alongside the theoretical development of the TSE scales, a large number of studies has been produced investigating correlates of TSE. Those correlates can be classified as those concerning characteristics of the teachers themselves, i.e. age, years of work experience or attitude towards teaching, classroom characteristics like student achievement, classroom climate or size, principal characteristics like leadership style and school characteristics like size or status or public or private. The following section will provide an overview of those studies that provide a relevant framework for the investigation of TSE in the 32 OECD countries and the three country clusters in chapter 6.

The predominantly used method of analysis were quantitative methods like classroom observations and questionnaires or a combination of them. These studies provide the researcher with information that has the potential to explore relationships or even predictions between certain observed traits. Thus, the following research is built upon a corpus of knowledge that can systematically be expanded.

2.3.1 TEACHER CHARACTERISTICS AND TEACHER'S SELF-EFFICACY

Teacher traits that have been considered are time related such as age or work experience. Chester & Beaudin (1996) found that TSE is mediated by age and prior work experience as in those studies, older and more experienced teachers had a higher sense of self-efficacy, which was supported by Cousins et al. (1996) and Fackler & Malmberg (2016). Chang (2008a) found a higher level for TSE in classroom management, student engagement and instruction for experienced teachers as compared to novice- or pre-service teachers; regardless of prior teaching experience. A non-linear relationship between TSE and work experience, in contrast, was found by Klassen & Chiu (2010), where TSE first rises until mid-career and then drops. A gain in TSE during pre-service teaching followed by a decline during the first year of regular teaching in relation to the support level was found by Burke &

Spero (2005). By contrast, no influence of work experience and academic level on TSE was found by Pas et al. (2012), but over the course of the career, TSE first appeared to rise and then dropped. Furthermore, using the tripartite measure of TSE comprising instruction, student engagement and classroom management, Wolters & Daugherty (2007) found that TSE in all three domains was positively related to teaching experience and could explain mastery goal structure. In addition to work experience, factors contributing to a high level of TSE reported by Shanghai teachers were the respect and confidence that teachers experienced both from students and parents and teacher training courses (Cheung, 2008). Taking into account the level of preparedness through teacher training and classroom experience, Giallo & Little (2003) found a positive relationship with TSE. Investigating the level of TSE in instruction, student engagement and classroom management for teachers with less, compared to more, than 15 years of work experience, Gkolia et al. (2014) found that less experienced teachers reported a lower level of TSE in classroom management, meaning that more experienced teachers felt more capable of managing their students' behaviour. The other two domains, however, showed no significant difference with respect to the teachers' work experience. Zee & Koomen (2016) synthesised that more experienced teachers might be more sensitive towards students' signals and needs and might thus be better prepared to adequately respond. However, these positive experiences do not seem to last over the course of the career.

For teacher gender, findings for the association with TSE are again mixed: While Cousins et al., (1996) found male teachers to have a higher TSE, Coladarci (1992) and Fackler & Malmberg (2016) found female teachers to be higher, but Malmberg et al. (2014) found no gender effect. Furthermore, Klassen & Chiu (2010) show that female teachers tend to report lower TSE in cases of high workload stress and stress with student behaviour. They seem to be more responsive to environmental factors than their male peers. A high workload was also related to a high level of TSE (Klassen & Chiu, 2010).

TSE was also found to have relationship with teacher attitudes and behaviours towards students and in the classroom. Emmer & Hickman (1991) found a positive relationship between TSE and positive strategies of student support. High TSE teachers were more likely to encourage students, introduce peer support, reflect on their own teaching, and use

more reductive strategies (e.g. desists, warnings), as well as more positive strategies, in order to achieve or maintain a desirable behaviour. The same was found for teachers who felt educationally well prepared and anticipated success in the classroom (Cousins et al., 1996; Raudenbush et al. 1992). Li & Liu (2000) could show that high TSE teachers tend to use more and more diverse academic activities, provide more in-depth feedback and use more cognitively challenging tasks. Also, more experienced teachers with a higher level of domain-specific TSE reported a higher level of mastery-experience in classroom situations (Malmberg et al. 2014). Teacher characteristics like creativity, analytical and practical skills were found to predict a higher level of TSE in all three domains of classroom management, student engagement and instructional strategies; particularly the latter ones (Chang, 2008b). Furthermore, perceptions of positive personality characteristics like communication skills or self-confidence were found to be related to a higher level of TSE in instruction and classroom management, whereas TSE in student engagement was found to be related to their motivation for teaching (Poulou, 2007).

2.3.2 SUMMARY

As can be seen from the review above, various teacher characteristics have been tested in their association with TSE. However, although they build on a long research tradition, only very few replications were done, and the choice of variables analysed together is versatile. Furthermore, only a comparatively small number of studies do not concern the Anglo-Saxon context but focus on i.e. Asian or Nordic teachers. This limitation (among others that will be identified later) will be addressed in this study. Especially the mixed effects found for gender or teaching experience hint at complex associations TSE has with teacher characteristics. It can only be speculated, however, that the differences could lie in the different instruments used and thus be evidence of inconsistencies in measuring the concept or other factors, which are difficult to capture retrospectively.

2.3.3 CLASSROOM CHARACTERISTICS AND TEACHER'S SELF-EFFICACY

The second group of characteristics are classroom traits and their association with TSE. Classroom characteristics in this context are those that emerge from the classroom composition and comprise mostly of student features e.g. their socio-economic background,

the classroom climate or student achievement. Due to the structure of the questionnaires used in TALIS (see Chapter 3 for a more detailed description of the data collection), this information is taken from the questionnaire that has been administered by the teachers and is thus based on the teachers' perception or knowledge about their class. This section gives an overview of past research on classroom characteristics of TSE and aims at highlighting its strengths and weaknesses as well as pointing out needs for further research.

One of the questions that have been addressed in research on the correlates of TSE is how student achievement and TSE are related. So by approaching TSE as a trait that students experience in the shift from a math teacher with a high level of TSE to a math teacher with a low level of TSE, the result was a decline of the pupils' self-perception of math-ability and increased difficulty of task (Midgley, Feldlaufer, & Eccles, 1989). The positive association of TSE with student achievement was supported by Mojavezi & Tamiz (2012) and Caprara et al. (2006), who found a positive relationship between TSE, student outcomes and self-determination (Alivernini & Lucidi, 2011) and print awareness for young children (Guo, Piasta, Justice, & Kaderavek, 2010). Additionally, if the classroom provides a supportive emotional climate, it was shown that there can be positive effects on vocabulary gains (ibid.). When individual coaching experiences for teaching improvement were taken into account, TSE as well as student achievement were higher (Ross, 1992). Teachers, who reported a high sense of TSE, turned out to also feel more responsible for their students' achievement than those with a low sense of TSE and also displayed a stronger belief in their own teaching abilities. Eberle (2012) as well as Vasquez (2008), in contrast, found no relationship between teachers' perceived self-efficacy and pupils' maths- and reading achievement.

Concerning the classroom setting, a bigger class-size and higher student age were predicting a higher level of TSE (Coladarci, 1992; Raudenbush et al., 1992). Klassen & Chiu (2010) found a relationship between three domains of self-efficacy (instructional strategies, classroom management and student engagement), two types of job stress (workload and classroom stress) job satisfaction and teacher socio-demographic information: female teachers had a lower sense of classroom management efficacy and a greater workload stress and stress from negative student behaviour. In contrast, teachers with a higher workload stress

had a greater sense of classroom management efficacy, whereas teachers with a high classroom stress reported a lower level of SE and also lower job-satisfaction. Furthermore, teachers with a higher level of TSE also reported a greater use of structuring, student-oriented and enhancing activities in the classroom and showed also a higher level of TSE with a larger number of students from an educated home, but also a different mother tongue from the language of instruction (Fackler & Malmberg, 2016). Also, in relation to workload stress, Klassen et al. (2013) in an international study found a positive relationship between all three domains of TSE and workload stress for English and Hong Kong teachers, but a negative relationship for Canadian and Thai teachers. Furthermore, TSE was found to mediate the relationship between stress in relation to student behaviour and job commitment for novice teachers. Dutch teachers of ethnic minority students were found to feel more self-efficacious than teachers of ethnic majority students in relation to classroom management; especially when there were more ethnic minority students in the classroom and problem behaviours were internalised. Hence, a more homogeneous classroom – both ways - seems to be supportive of a higher level of TSE. However, a lower level of TSE for minority students was reported in relation to instruction and student engagement (Geerlings et al., 2018). In relation to ethnically diverse classrooms, Tucker et al. (2010) described a successful teacher workshop and follow-up consultations that had arisen awareness and sensitivity for the multiple social, cultural, political and family challenges those children face and was able to increase the teachers' cultural TSE. Another result shown by this study was increased confidence in the management of diverse classrooms.

Furthermore, Rimm-Kaufman & Sawyer (2004) reported that high TSE teachers are more likely to introduce new teaching approaches and stricter classroom management. For South Korean teachers, a high level of TSE was associated with positive interpersonal experiences like a good classroom climate as well as job satisfaction and good student-teacher relations (Aslan, 2015). High TSE teachers also use more instructional teacher-centred teaching strategies (Klassen & Chiu, 2010) and are more likely to create a good classroom atmosphere (Ashton et al., 1983). This is supported by Künsting et al. (2016), who reported that self-efficacy is associated with a good classroom climate, the use of cognitive

activation strategies and successful classroom management strategies. They could furthermore show that this trait is stable over time.

Another classroom trait that has been investigated repeatedly with respect to its association with TSE is student achievement and engagement in the lessons. A positive association with TSE and students' achievement was i.e. reported by Malmberg et al. (2014), Fackler & Malmberg (2016) and Ashton et al. (1983). Similarly, the transition from a high to a low TSE teacher appeared to have a negative association with the students' math achievement (Midgley et al., 1989). A reciprocal relationship was found for the students' engagement in lessons: it accounts for a high level of TSE (Guo et al., 2011; Pas et al., 2012) and, vice versa, high TSE teachers were found to better engage students in the lessons (Holzberger et al., 2014).

2.3.4 SUMMARY

The review shows that much research has been conducted to explore the influence of classroom characteristics on TSE with respect to features like teaching strategies, student achievement or classroom management and atmosphere, but, as it has also been the case for the teacher characteristics. most of the research was conducted in the North-American context. Only, for instance, the very recent studies by Künsting et al., 2016 and Geerlings et al., 2018 investigate an international, respectively German or Dutch classroom context and none explicitly seek to explore the possible role of cultural differences. Thus, there is a lack of intercultural knowledge on how classroom characteristics may affect the development of TSE, which is one of the aims of this research.

2.3.5 PRINCIPAL AND SCHOOL CHARACTERISTICS AND TEACHER'S SELF-EFFICACY

The last suggested associations with TSE concern the principal and the school environment. They are deemed important as they partly form the conditions in which teachers work; the principal as the head of school and responsible person for decision making, the school environment as the broader social determinants of the teacher's work place.

Investigating school environmental factors, Coladarci et al. (1992) found that TSE was positively related to a favourable school climate, a smaller student-teacher ratio as well as a principal's instructional style of leadership. This is supported by Fackler & Malmberg (2016), who could also show that an experienced principal has teachers with higher levels of TSE as well as Guo et al. (2011), Raudenbush et al. (1992) and Chester & Beaudin (1996). They could also show that staff collaboration is positively connected to TSE. Furthermore, a supportive and academically accomplished principal who uses connections with authorities in the teachers' favour and as trust and respectful relationship contributes to a strong sense in general TSE (W. Hoy & Woolfolk, 1993; Skaalvik & Skaalvik, 2007). Thus, it seems that teachers feel more self-efficacious in an environment that provides them with a clear goal-structure and an atmosphere of appreciation. This indicates that strong leadership is an important factor in shaping a positive work environment for teachers. In a meta-analysis on several principal leadership styles, Kirk (2016) found the strongest association for teacher's self-efficacy and an instructional leadership style. Furthermore, the opportunity for teachers to influence decision making (Ashton et al., 1983), as well as a strong emphasis on academic achievement (W. Hoy & Woolfolk, 1993) have also been linked with high TSE.

Not only principal related factors, but also cooperative parental relationships showed a significant association with TSE (Skaalvik & Skaalvik, 2007). In contrast, student mobility, school enrolment, student suspension rate, free and reduced meals, principal turnover (%) and collegial style of leadership does not seem to be associated with TSE (Pas et al., 2012). Hipp (1996) distinguished between a transformational style of leadership (the sense that "teachers in general can accomplish things" as defined by Guskey & Passaro, 1994), which has an impact on general TSE, whereas role model behaviour of the principal turned out to be positively related to personal TSE. This was supported by Walker & Slear (2011), who additionally accounted for the principal's work experience. Furthermore, the mediating role of TSE between an instructional style of leadership and collective self-efficacy was pointed out by Calik et al. (2012), who could show that both features of TSE were positively related to an instructional style of leadership.

Considering the environment of schools, Knoblauch & Woolfolk-Hoy (2008) showed that TSE increases over time, but less so for teachers who teach in urban areas compared to those who teach in rural or sub-urban ones. Moreover, teachers in private schools reported a higher level of TSE than their colleagues in public schools (Fackler & Malmberg, 2016). Recently, Holzberger et al. (2014) found that in schools that cater for their teachers' intrinsic needs (i.e. autonomy or competence) teachers had higher TSE than in those where this was not catered for. On behalf of the opportunities the school can offer a teacher, Mizell et al. (2014) argue for a strong system of professional development in order to strengthen teachers' beliefs in their own teaching and personal skills. This was supported by Pinnock (2006), who could show that professional development for science teachers enhances their sense of self-confidence and teaching-efficacy. School features like a supportive principal that generates a positive school climate and promotes cooperation among teachers, good relationships with students and families as well as colleagues were found to be contributing to a higher level of TSE; but also to job satisfaction (Caprara et al., 2003). Finally, Ross & Bruce (1997) showed that professional development positively influences teachers' beliefs in classroom-management efficacy.

2.3.6 SUMMARY

As can be seen from the review above, in past studies attention has been paid to the principal in TSE studies; in particular the leadership style and goal-setting, but also development opportunities the teachers are offered at school. No attention at all seems to have been paid to the person of the principal as such by considering i.e. their age, gender or work experience. Less systematically and frequently, other school characteristics like size, status being public or private or school location were investigated. Thus, a comprehensive investigation of these factors together with the teacher and classroom characteristics described in previous sections to take account of the multi-faceted environment in which teaching takes place has not yet been conducted. This might be due to the lack of large-scale studies that include the perspective from both, teachers and principals and also collect information about the classroom and school, which provide contextual information about the teacher's work environment. In the proposed research I will address these gaps using the TALIS data set, which provides a wide range of those factors.

2.3.7 COUNTRY CHARACTERISTICS OF AND CULTURAL VALUES RELATED TO TSE

As noted earlier, one of the main features of TSE so far is that it has mostly been investigated in the North-American context (Klassen et al., 2010). Only in recent years, researchers in the Asia-Pacific countries have begun to research TSE in this region, but have mainly been concerned with measurement issues and the validity of the instruments. Also, some of the publications are not available in English, which limits connecting this research internationally (Malinen, 2016). Other countries in which TSE has been explored include the Nordic countries, where Skaalvik & Skaalvik (i.e. 2007, 2010, 2014) did some seminal work. This, on the one hand, limits the generalisability of the results towards other countries and cultural contexts. On the other hand, there has been no study so far that allows comparison of different education systems, reflecting a lack of data to do so. Drawing on intercultural literature and applying those concepts to the education context can address this gap. Furthermore, national indicators that have been used in other contexts such as in the PISA study might help to shed light on the influence of the wider context in which TSE develops, as Vieluf, Kunter, & Vijver (2013) suggest.

Thus, this study will have two parts: In the first part, predictors of teacher self-efficacy using the tripartite concept suggested by Tschannen-Moran & Hoy (2001) will be analysed across 32 OECD countries using data from the TALIS study. In the second part, three country clusters (comprising Anglo-Saxon, Nordic and East- and South-East Asian countries) will be compared as to how similar and different those predictors relate to TSE. With this, it is hoped to shed light on cultural differences in which teaching takes place and how the different work environments in which teachers find themselves across the globe shape the teachers' sense of their own self-efficacy.

The following model shows how TSE is understood in its relation to the different predictors, which illustrates the way TSE will be investigated in this thesis.

Fig. 1 1 Own model of characteristics of TSE

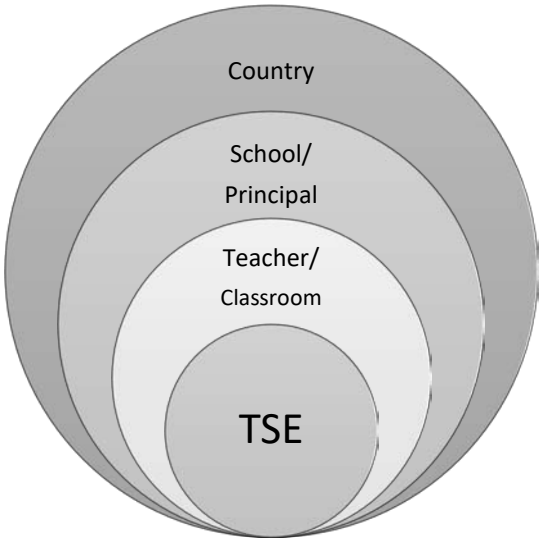


Figure 11 shows how TSE is perceived in this study in terms of nested levels:

TSE as the key concept is in the very centre of the study with the predictors surrounding it considering the hierarchy of the data structure and closeness to the teacher from proximal to distal: The first layer represents the teachers themselves and the classroom, the second layer the school/principal and the third and outer layer the country in which the teacher teaches. The country level, however, serves as an interpretive framework to explain differences in the teacher's self-efficacy in the country clusters mentioned above. It is not possible to further distinguish between the teacher/classroom and principal/school level, as teachers report on their classes and principals on their schools.

CHAPTER 3 METHODOLOGY

The aim of this chapter is to explain the methodological approach to answer the research questions. First, ethical considerations of secondary data analysis will be explored, as well as a brief section on its epistemological underpinnings. Then there is a general section that describes the data sources and variable treatment that is relevant to both research questions. Methodological considerations for research question two will be addressed in a separate chapter later in the study.

3.1 ETHICAL CONSIDERATIONS

As required by the ethics committee of the University of Oxford, ethical approval of the project was obtained. In this particular case of the analysis of international secondary data, there were two steps: First, ethical clearance was approved by the single participating countries before the data collection started. The OECD and the GLOBE researchers, respectively, their national project managers in each country, made sure that the data collection, storage and treatment followed the requirements of each country. This includes the full anonymisation of each questionnaire, so that answers cannot be traced back to participants. Second, this has been counterchecked and approved by the CUREC process of the University of Oxford. Accordingly, the risk of possible harm for the study's participants can be assumed to be minimal. The letter of approval for CUREC is in Appendix 1.

The OECD data (PISA and TALIS) are freely available on the website of the OECD for research purposes (http://stats.oecd.org/Index.aspx?DataSetCode=TALIS; http://pisa2012.acer.edu.au). In the same way, the instruments were made available to the researcher. The data of the GLOBE-study will be taken from the publication itself (House, R., Hanges, P., Javidan, M., Dorman, P., & Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. SAGE Publications, Inc.).

3.2 SECONDARY DATA ANALYSIS

The present study is an example of quantitative research that seeks to explore patterns and associations and to investigate and test hypotheses about teacher self-efficacy. It is a secondary data analysis – secondary for the reason that the data has been collected by a different institution than that to which the analysing researcher belongs and for a different original research purpose than the present one. Such data sets often come from governments, national or international organisations whose resources allow large-scale collection which is often not available for a single researcher or university. Such organisations typically also have the resources to employ sophisticated data collection methods and experienced researchers to gain high-quality data (Vartanian, 2011). From this, however, several advantages and disadvantages emerge.

3.2.1 EPISTEMOLOGICAL FOUNDATIONS OF QUANTITATIVE SECONDARY DATA ANALYSIS

Quantitative social science/education research is traditionally based within the post-positivist framework of epistemology (here and in the following: Bryman, 2010 & Pring, 2004). Epistemology as such states what can be deemed acceptable knowledge in a discipline. Post-positivism, however, is a further development of the positivist research, which can be described as the systematic observation of what '*is clear, factual and open to observation*' (Pring, 2004; p. 91). Key principles of the acquisition of knowledge in the positivist framework can be described as the following (Bryman, 2010):

1. Knowledge can only be gained through observation of the environment. This can be done by using a variety of research instruments, i.e. questionnaires.

2. Research is based on falsifiable hypotheses, which are generated based on theory. They provide the basis for general laws of explanation, which in turn can inform and develop theory.

3. Science should be conducted in a way that is value free and transparent so that it can be critiqued.

4. One should distinguish between normative and scientific statements. Scientific statements can be falsified by the data, whereas the truth of a normative statement is beyond the possibility of investigation of the researcher.

The post-positivist approach acknowledges some of the limitations of this approach, the strict principles of which hardly hold in practice; for example, in much of the world that which the education researcher is interested in investigating cannot be directly observed; examples relevant to this enquiry include teachers' attitudes or beliefs. These multifaceted concepts, then, are more driven by theory than mere observation. It is also not often possible to exclude the normative part that can be hidden within a scientific statement: the choice of topics, instruments and participants might, although well justified, still be influences by underlying normative assumptions. In this study, however, the aim is to on the one hand to work as rigorously as possible and on the other hand make these issues transparent where possible and thus follow a more defensible post-positivist approach (c.f. Crotty, 2009). This allows the acknowledgement of the subjective element that remains, regardless of rigorous methods and is a more tentative or even humble and realistic approach towards research results.

3.2.2 Advantages of Secondary data analysis

One advantage of secondary data analysis is that the financial and personnel resources of big organisations like the OECD allow for a scope of the project which is difficult to achieve for a single researcher (here and in the following: Smith, 2008). Especially in education, where large-scale international comparative analyses like PISA (Programme for International Student Assessment), TIMMS (Trends in International Mathematics and Science) and TALIS (Teaching And Learning International Survey) are becoming commonly used, these data sources promise to give useful insight and inform education research, politics and decision making in general. Due to the large size and careful sample frames adopted, these data sets also allow researchers to draw generalisable conclusions, as they are often representative of the population. Furthermore, due to their statistical properties, these large-scale data sets like TALIS allow for the use and development of complex and sophisticated

statistical techniques. Thus, analyses of such secondary data sets not only contribute to knowledge gain, but also to methodological development in education research.

3.2.3 DISADVANTAGES OF SECONDARY DATA ANALYSIS

Despite these promises of secondary data analysis, there are also some disadvantages or limitations (here and in the following: Smith 2008; Vartanian, 2011). The first one is that the researcher has only access to the variables present in the data set and might need to adjust their research accordingly. The second is that the choice of indicators might be biased by the aims of the organisation collecting the data. With respect to the OECD, for example, this could mean an economic focus on the scope of the study. As described previously, the OECD views education as a means to economic ends, which serves as an umbrella for all its studies into the operation and features of different education systems. These possible biases have to be taken into account when drawing conclusions or attempts at generalisations from studies based on such secondary data. The third limitation to be mentioned here is that these data often take years from being collected to being published, so such data might not be the most recent information, but rather the most recent that is available (Smith, 2008). Fourth, there is a lack of control: All information about the data collection process, data quality control and management done prior to making the data accessible as a secondary data set is beyond the control of the analysing researcher. Accordingly, there might be 'blind spots' one cannot refer to with respect to the data collection process and their implications for the results – both, in positive and negative ways. For the TALIS study, however, the OECD has published three major resources:

1. *The Technical Report (OECD, 2013)*: The 'Technical Report' reports on the development and implementation of the study. It provides information on the theoretical framework, sampling strategies and decisions, instrument development and implementation, data management and the construction of scales and indices.

2. *The principal and teacher questionnaire (OECD, 2013)*: The full questionnaires, as they were given to the study participants, are made available on the internet (https://www.oecd.org/edu/school/talis.htm).

3. *The Conceptual Framework (Rutkowski et al., 2013)*: The Conceptual Framework provides background information on the rationale of the research, its general purpose and policy relevance, chosen themes and main indicators as well as the study design and implementation.

4. *Various related resources*: The OECD itself publishes results and papers on the TALIS data set (see https://www.oecd.org/edu/school/talis.htm for a list of publications using TALIS data).

3.3 THE OECD, EDUCATION AND TALIS

The following overview aims firstly at providing a summary of the work of the OECD on education. It attempts to show the relationship between the various efforts on investigating the education system along the life span and thereby put TALIS in perspective to the work of the OECD in general. Secondly, it provides in-depth information on the TALIS study and its methodology.

3.3.1 THE OECD AND EDUCATION

The 'Teaching And Learning International Survey' (TALIS) is part of the research in education the OECD conducts. In order to promote policies that support economic growth and provide a basis for prosperity (OECD, 2019), the OECD initiates many comparative studies in the education systems of its member countries along the span of life. Under the umbrella of the INES (INdicators of Education Systems) project, a myriad of studies and networks exist, which investigate the quality of the education systems of the member countries and their policies (here and in the following: OECD 2012): Encouraging Quality in Early Childhood Education and Care (ECEC) aims at finding and implementing successful policies that enable 'equitable access to quality early childhood education and care' (OECD 2012, p. 5). The school system is investigated by two separate studies (which are linkable for some of the countries): The PISA (Programme of International Student Assessment from 2000) and TALIS (Teaching and Learning International Survey from 2008) studies. While PISA focuses on transferable skills of the students, which should help them on the labour market, and the ability of the school systems to overcome social inequality, TALIS investigates the

school environment in which teaching and learning are taking place. Teachers and principals report on their attitude towards various aspects such as features of the school environment, teaching and learning, collaboration and professional development and their work in the classroom. Principals report also on their school leadership and work environment. In the tertiary sector, the Vocational Education and Training (VET) project examines how education systems deliver necessary skills for non-academic job training. The Assessment of Higher Education Learning Outcomes (AHELO) aims at figuring out factors that contribute to a successful higher education system in the respective countries by investigating areas like system objectives, quality assurance, finance structures, policies and governance mechanisms. The last area to be mentioned here is the Programme for the International Assessment of Adult Competencies (PIAAC) that aims at investigating and fostering perspectives and policies in adult learning; both in- and outside the workplace.

3.3.2 THE TALIS STUDY

The TALIS study in particular investigates the environment in which teaching and learning are taking place. It aims at providing information that is relevant for policy makers. It attempts to provide value added estimates, and indicator oriented, rigorous, valid, reliable and comparable information that is interpretable and cost-effective to let the participating countries benefit from it (Rutkowski, 2013) (for more in-depth information on the theoretical framework, see Sections I & II). The following section describes the compilation of the questionnaire, sampling and data collection strategy.

3.3.3 COMPILATION OF THE QUESTIONNAIRE

The compilation of the questionnaire was mainly decided upon by the participating countries; but jointly with the OECD responsible people. Hereby, the OECD aimed at following a fair approach by giving the participating countries equal votes to rate their preferred themes that should be followed in the study. There were two requirements: First, as TALIS is designed to be conducted every five years, countries were asked to identify themes that should be followed up from the first sweep. Second, countries were asked to determine which additional themes and indicators to include in the second sweep. These would then result a teacher and principal questionnaire in each country, which should take between

45-60 minutes to complete. The rating required 200 scores to be distributed among 20 themes; higher scores indicating higher priority, included 94 indicators. In all, 25 of the indicators had to be identical with those of the first sweep to follow up changes from survey 1 to survey 2. In total, 25 countries out of the 34 took part in the vote. As only those countries, which had already been part of the first round of TALIS, were allowed to vote on the repeated scales, Belgium, France and the United Kingdom did not report on the repeated indicators. Nevertheless, they still got to vote on the other scales. The main themes and indicators identified here were: School leadership, teacher training and in-service professional development/initial teacher education, teacher appraisal and feedback, school climate and ethos, teachers' pedagogical beliefs and teachers' pedagogical practices. These were put in practice using established research instruments.

After deciding upon the themes and structure of the questionnaire, a pilot study was carried out to inspect the length and quality of the instruments before the final questionnaires were drafted with the remaining items and scales. Furthermore, all items and questions were translated into the national languages, but a reference questionnaire in English was available for both teachers and principals (Rutkowski et al., 2013).

3.3.4 SAMPLING

The OECD defines teachers as persons, "*whose professional activity involves the planning, organising and conducting of group activities whereby students' knowledge, skills and attitudes develop as stipulated by educational programmes. In short, it is one whose main activity is teaching"* (OECD, 2004, p.47; in: OECD 2013, p. 72). To keep the focus of the study clear, the possible population only comprises of all teachers in a country, who teach students at ISCED level 2; teachers teaching children with special needs were only included in the sample if they taught at regular schools. The principals of those schools were also included. Countries were given the opportunity to exclude from the sample those schools that had fewer than three teachers teaching ISCED 2 and in remote areas, for economic reasons, or only select political entities (i.e. administrative regions) if they would still include at least 95% of the national survey population. Thus, the national target population and survey population could differ for those countries.

In order to provide a sufficient sample size for generalizable results, a two-stage clustered sampling strategy was applied and at least 200 schools with 20 ISCED 2 teachers were randomly selected from the survey population. Accordingly, at least 4000 teachers and 200 principals from each country were included. Teachers working in more than one school were included in the sample, but adjusted for the number of schools (i.e. a teacher working in two schools were only counted as half a teacher in each school). In contrast, those teachers, who also worked as principals were excluded from the sample, as well as those who were on long-term leave. Countries were given the authority to increase both the sample size of teachers and schools, and for a few countries that had less than the required number of schools, these were adjusted, too (see OECD 2013, Annex E, for greater detail). Moreover, countries had to define strata into which schools were classified, from which a random sample of schools was drawn. If any of the selected schools opted not to participate, a school above or below was chosen as it was assumed it has similar features and provide sufficiently accurate estimates compared to the one originally chosen. This was carried out by Statistics Canada and national project managers (NPM), who were trained in the use of the guiding project manager material, were employed to liaise with the national school authorities, schools and the OECD. Within each school, a school coordinator was appointed to oversee data collection and to be contact person for the NPM (OECD, 2013).

3.3.5 DATA COLLECTION

The questionnaires compiled by the countries were handed to the teachers and principals either as pen and paper or electronic versions. If teachers or principals decided to fill in the paper form, the data was added to the data base by responsible statisticians using the IEA data management software. To keep the survey manageable, countries were assigned to either the Northern or Southern hemisphere, depending on which side of the equator they were located. The data collection took place in each country towards the end of the school year to assure comparability and lasted between 12 days to four months. Although 100% participation rate within each school was aimed at, a school was included in the study when at least 50% of the sampled teachers responded. For the schools, a response rate of 75% was aimed at. This was achieved for all countries.

For confidentiality purposes, all data was anonymised and schools/principals and teachers were assigned unique identifiers that allowed to follow the nesting structure of teachers in schools, but not to identify the persons behind it. In the tracking forms, however, the names of the respective participant were still listed, but these were only accessed by the NPMs and fully anonymised before publication of the data. It was made sure that these procedures strictly follow national data protection regulations.

For the OECD to be informed of problems or difficulties throughout the data collection process, a survey activities questionnaire was to be filled in by the project managers. This also helped the organizers to be aware of limitations and follow up reasons if the survey fell behind standards that were defined prior to data collection; as, for example, in the case of one country that provided fewer teachers within schools than was aimed at. Another set of tools made available to the NPMs were handbooks and frameworks that provided background information on the development and planning of the survey (whole section: c.f. OECD, 2013).

3.3.6 IMPLICATIONS FOR THE USE OF OECD DATA

It should nonetheless be pointed out that the use of OECD data might have implications beyond the intended scope and therefore some points of critique should be mentioned, too. It should also be noted here that most concern regarding OECD studies and their impact on education policy has been directed towards PISA, which is the most well-known of their studies into the education systems of youth's reading comprehension, maths and science problem-solving skills. Thus, much of the critique mentioned here is made by making inferences from PISA to TALIS. This seems to be appropriate where the critique addresses topics that concern the description of the education system (i.e. direct measures of school or teacher characteristics or where more general psychometric scales are used to assess certain attitudes like beliefs about teaching or the leadership style). Critique of PISA that, in contrast, addresses, for example, the use of international assessment data is too PISA-specific and therefore it is not deemed appropriate to extend it to the TALIS study.

Although the OECD encourages the use of its data for making evidence-based decisions in education policy and points out that it uses strict academic approaches when carrying out

its studies, the data should nonetheless only be used cautiously (Hopfenbeck et al., 2017): The OECD studies often are the only point of references for countries to investigate the state of their education systems, as, for example, only the USA or England have a longer tradition of collecting large-scale data on schools and students. Thus, topics that are enhanced by the OECD might not necessarily overlap with the topics of original interest to the countries. Therefore, it might shed only partial light on areas of interest for single countries and lead to policy decisions that are not exactly tailored towards the country's needs. Furthermore, the fact that the OECD is an internationally operating institution that appears with power, might tempt countries with only weak education governance to only use this data as justification for policy changes. This means the OECD 'sneaks in through the back door' and influences decisions that are incumbent upon the sovereignty of the countries. Furthermore, it can be said that for some countries (i.e. Norway, see Moeller & Skedsmo, 2013), the desire for data driven large-scale analysis only emerged after the OECD had implemented PISA and for those countries created a demand for these. TALIS can be seen as project emerging from that by extending the investigation in the education systems to teachers and schools. Another strand of concerns is targeted towards the OECD as an economic organisation, which sees education only as a means to achieve economic ends, but not as an end in itself (Sjöberg, 2015). Thus, the view the OECD has on education is somewhat limited and not even fully endorsed by businesses and companies, which also requires their employees to have soft skills and a moral compass, aspects the OECD does not include in their studies (ibid.).

In summary, it can be said that although the OECD is providing a rich data source into the school systems of its member countries, the results emerging from it should be treated cautiously. Furthermore, the data should ideally only be used to serve the specific interests of the respective country. The ranking of countries or comparisons do not seem to do justice to the very different education systems and states in which these are. And, finally, the OECD, as such, is a think tank, but not an official policy institution. Thus, the data it provides can at best be used to inform education policy where of interest.

3.3.7 SUMMARY

This section aimed at providing an overview of the background of the theory and practice behind the set-up of the TALIS study. It shows (1) that the decisions about the content of the study were jointly made by the participating countries and thus were made to cater to the perceived needs of their education systems, (2) that the data collection and management till the publication of the data was well taken care of and handled in a professional way and (3) that both ethical and research standards were respected. Although this provides promising secondary data for the intended research purposes it is nonetheless important to point out that the researcher is reliant on the quality of the data provided and has to use the documentation provided and make a careful study of distributions, patterns of missingness and descriptive statistics to establish confidence in the robustness of variables and constructs and their applicability in the context of different country. (4) The conceptual caveats described above should always be borne in mind. The role of the OECD as a think tank that provides education policy makers with information has long been scrutinised and it is important to consider the impact it can potentially have and be aware of its limitations.

3.4 ANALYTIC PROCEDURE

The following section of this chapter aims at providing an overview of the analytical procedure employed and explaining the choice of methods. It explains the need for multilevel analysis rather than a standard regression analysis, to answer the first research question. Then it points out possible sources of error, describes types of variables that occur in multilevel data, explains the need for centering variables, explains how to disentangle variation across levels and the set-up of the statistical models adopted to address the research questions.

3.4.1 MULTILEVEL ANALYSIS

Data in the context of the present study is hierarchical, reflecting the existence of clusters in the school setting (for example, teachers nested in schools nested in countries). How-

ever, it is important to note that these clusters or 'levels' are constructed by i.e. the political or geographical context in which they occur. Because of these relationships between the entities of which these pre-existing clusters consist, the assumptions of independent observations and identically distributed variables, which are necessary for the traditional regression approach, will not be met. Teachers in the same school may share certain characteristics with each other, which teachers from different schools do not share. The same is likely to be true for schools from the same country compared with schools located in different countries. Accordingly, because of the structure of the data and the particular interest of comparing effects internationally, multilevel modelling is deemed to be the appropriate tool of analysis. (Goldstein et al., 1995; Raudenbush & Bryk 2002; Hox, 2010).

1. *The number of observations on each level should be sufficiently large*. Mass & Hox (2005) and Hox (2010) have shown that for achieving robust variance components, there should be at least 30 observations per level. Furthermore, taking into account higher level observations can provide helpful further information and the authors argue for a relation of 50/20 between the higher and lower level observations or even 100/10, if the variance components are of particular interest. Accordingly, the ratio of observations between each cluster of the higher and lower level should be greater than 2.5. The descriptive statistics show that the TALIS data fulfils these requirements: As shown in Table 1, the ratio between teachers/school was between 9.416 – 20.559 and schools /countries was 122-1070.

2. *There should be no multicollinearity in the data*. Multicollinearity requires the independent variables being independent from each other and the correlation coefficient for each pair of variables should be < .9 (Tabachnick & Fidell, 2007). This was tested for by correlating the independent variables of each level with each other. In this research involving the TALIS data set no correlation was found to be higher than $r = .8$; $p < .001$ and thus multicollinearity can be excluded. See Appendix 4 for the full correlation matrix.

3. *The level-one-data should be homoscedastic and normally distributed*. When the data are homoscedastic, the variances of the observations are equal. If this assumption is violated, the standard errors of the data are biased. A PP-plot check (see Appendix 5.5) for the level 1-variables revealed that normality and homoscedasticity can be assumed, as there are no significant deviations from the normal line.

Table 1 Descriptive statistics for teachers, schools and countries

Country	n Teachers	n Schools	n Countries	T/S	S/C
Australia	2,059	123	1	16.740	123
Brazil	14,291	1,070	1	13.356	1070
Bulgaria	2,975	197	1	15.102	197
Chile	1,676	178	1	9.416	178
Croatia	3,675	199	1	18.467	199
Czech Republic	3,219	220	1	14.632	220
Denmark	1,649	148	1	11.142	148
Estonia	3,129	197	1	15.883	197
Finland	2,739	146	1	18.760	146
France	3,002	204	1	14.716	204
Israel	3,403	195	1	17.451	195
Italy	3,337	194	1	17.201	194
Japan	3,484	192	1	18.146	192
Korea	2,933	177	1	16.571	177
Latvia	2,126	116	1	18.328	116
Malaysia	2,984	150	1	19.893	150
Mexico	3,138	187	1	16.781	187
Netherlands	1,912	127	1	15.055	127
Norway	2,981	145	1	20.559	145
Poland	3,858	195	1	19.785	195
Portugal	3,628	185	1	19.611	185
Romania	3,286	197	1	16.680	197
Serbia	3,827	191	1	20.037	191
Singapore	3,109	159	1	19.553	159
Slovak Republic	3,493	193	1	18.098	193
Spain	3,339	192	1	17.391	192
Sweden	3,319	186	1	17.844	186
United States	1,926	122	1	15.787	122
Sub-national entities					
Abu Dhabi (United Arab Emirates)	2,433	166	1	14.657	166
Alberta (Canada)	1,772	182	1	9.736	182

England (United Kingdom)	2,496	154	1	16.208	154
Flanders (Belgium)	3,129	168	1	18.625	168

* T/S: Ratio of teachers/school; S/C: Ratio of schools/country

The data particularities described in Table 1 support the use of multilevel modelling (MLM) (Tabachnick & Fidell, 2007). MLM allows the researcher to define variables on the level on which they occur and thus enables statistical modelling of hierarchical data. Hereby, one can distinguish between the micro, - meso- and macro level; starting with the individual (teachers here) at the micro level, going up to schools (meso level) and countries (macro level) (Snijders & Bosker, 2011).

3.4.2 SOURCES OF ERROR

But, as Hox (2002) points out, defining variables on different levels could lead to different interpretational errors, if the multi-level structure is not thoroughly followed through. One mainly distinguishes between the atomistic and ecological fallacy. The atomistic fallacy means to interpret individual level results at a higher level, whereas the ecological fallacy means to attribute group level results to the individual (Hox, 2002). It is nonetheless possible to aggregate data on the higher level of analysis, by, for example, calculating the mean of teachers' self-efficacy. This would be an indicator of the average confidence teachers have in their own teaching at a particular school. Second, any relationships between, for example, the age of the head teacher and TSE at school level, should, again, not be extended to the single teacher, as it is distinctly different from TSE at the individual level. The third source of error is that information might get lost if data are only investigated on the aggregate level, as the differences between individuals in a group are neglected (Snijders & Bosker, 2011).

3.4.3 TYPES OF VARIABLES IN MULTI-LEVEL DATA

Hox (2002) distinguishes between three types of variables that can occur in multilevel research, depending on the level on which they occur and refer to. *Global* variables only provide information on the level on which they are defined (i.e. teacher and classroom variables at the respective level as well as principal and school variables), whereas *structural*

and *analytical* variables, in contrast, provide information about the sub-units at a lower level. This type is constructed by aggregating the mean of a lower-level variable, for example teachers' age or work experience, which gives information about the average age or work experience of teachers at a certain school. The third type are *contextual* variables, which are values of a higher-level variable assigned to all lower level variables; i.e. whether the school is public or private. Here, information is disaggregated and provides information on the context in which lower level characteristics occur. Hox has pointed out, however, that the classification of variables into global, structural and contextual variables does not make a difference to the analysis itself but provides structure and guidance to the interpretation of the variables.

3.4.4 CENTERING VARIABLES

One possibility to approach the dependency of variables in multilevel research is to centre them around either the grand- or the group mean. In the present study, all continuous variables were centred around the grand mean by subtracting the grand mean from every individual observation, as the relationship between the dependent variable and level 2 predictors were of interest (Enders & Tofighi, 2007).

Equation 1 Grand-mean centering

$$x_{GMC} = x_{ij} - \bar{x}$$

with:

x = deviation score
x_{ij} = observed value of x
$\bar{x}$ = grand mean

(c.f. Raudenbush & Bryk, 2002).

Centering around the grand mean changes the raw score of the variable into a deviation score, but does not change the residuals or the model fit (Kreft et al., 1995; Paccagnella, 2006; Tabachnick & Fidell, 2007). Additionally, this alleviates the correlation between the intercept and slope estimates across groups, which reduces multicollinearity between level 2 predictors and so yields more reliable estimates (Raudenbush & Bryk, 2002; Kreft et al.

1995). Moreover, the meaning of the intercept changes: When centering the IVs around the grand mean, the intercept is interpreted as the expected value of the DV when all explanatory variables are equal to their mean value, and the higher-level intercept variance becomes the variance at the mean of all IVs (Hox, 2002; Raudenbush & Bryk, 2002). As there are only main effects tested in the models, this will not affect the interpretation of the model parameters (ibid.).

Group mean centering was not deemed appropriate, as the intention was to reduce the multicollinearity between higher and lower level variables, which is achieved by grand mean centering.

3.4.5 THE INTRA-CLASS CORRELATION

In MLM it is of particular interest to disentangle the relationship between the levels on which observations occur (Hox, 1995). First, the variance is partitioned into its components at different levels (such as teacher, school and country in a three-level model, for example), from which the intra-class correlation (ICC) is calculated. The ICC is a measure of how much variation on one level is explained by the overall variation in the model. Following the notation proposed by Hox (1995), the variance attributed to each level (or intra class correlation), was calculated by dividing the variance on each level by the total variance:

ICC on the teacher level:

Equation 2 Intra-class correlation on the teacher level

$$\frac{\sigma_t^2}{\sigma_t^2 + \sigma_s^2 + \sigma_c^2}$$

ICC on the school level:

Equation 3 Intra-class correlation on the school level

$$\frac{\sigma_s^2}{\sigma_t^2 + \sigma_s^2 + \sigma_c^2}$$

ICC on the country level:

Equation 4 Intra-class correlation country-level

$$\frac{\sigma_c^2}{\sigma_t^2 + \sigma_s^2 + \sigma_c^2}$$

with:

σ_t^2 = variance between teachers

σ_s^2 = variance between schools

σ_c^2 = variance between countries

The ICC itself is the proportion of total variance attributed to membership of different levels. The ICC was calculated for the null models first, and then again after predictors were added to the empty model. This strategy shows how far the inclusion of variables accounts for variance at each level and the variance that remains unexplained at each level. In the present study, the highest ICC occurred between the teacher- and all other levels (between 93.6%-99.6% across all subscales and models), the lowest between the schools- (between 0.4%-2.4%). The explained variance between countries ranged from 0.0%-6.4% across all subscales and models.

3.4.6 Analytical procedure: How is TSE in TALIS 2013 predicted by teacher characteristics, classroom characteristics and characteristics of the school and principal?

In total, the analysis of the first research question contains five models. The variables were entered in the analysis in the following order: In the baseline/null model ('Model 0') there are no explanatory variables, whereas in the following ones, explanatory variables are stepwise added to the analysis. In my research using the TALIS data set, the first model only contains teacher related variables, the second model adds in classroom characteristics, the third model includes principal and the fourth model then adds in school characteristics. The variables in the first two models operate on the first/teacher level, the variables in the third and fourth model on the second/school level, which is represented by the principal. For the third level (countries), there are, however, no explanatory variables available.

Overall, the research question comprises of four sets of analyses: The first analysis is a three-level model, which investigates the variation of TSE across all 32 countries. The second analysis consists of a two-level model, which is run in three country clusters with Anglo-Saxon, Nordic and East- and South-East Asian countries to investigate in more detail where the differences shown in the three-level model may appear. Chapters 5 and 6 will give an account of the theoretical framework and methodology used to investigate this.

How is TSE in TALIS 2013 predicted by teacher characteristics, classroom characteristics and characteristics of the school and principal?

Formally, research question 1 (cross-country analysis) is stated as follows:

1. Step: Teacher predictors

Equation 5 Model 1 RQ1: Teacher predictors

$$TSE_{ijk} = \beta_{0ijk} + \beta_{BLOCK1} Teacher\ characteristics_{ijk} + u_{0k} + u_{0jk} + e_{ijk}$$

2. Step: Classroom predictors

Equation 6 Model 2 RQ1: Classroom predictors

$$TSE_{ijk} = \beta_{0ijk} + \beta_{BLOCK1} Teacher\ characteristics_{ijk} + \beta_{BLOCK2} Classroom\ characteristics_{ijk} + u_{0k} + u_{0jk} + e_{ijk}$$

3. Step: Principal predictors

Equation 7 Model 3 RQ1: Principal predictors

$$TSE_{ijk} = \beta_{0ijk} + \beta_{BLOCK1} Teacher\ characteristics_{ijk} + \beta_{BLOCK2} Classroom\ characteristics_{ijk} + \beta_{BLOCK3} Principal\ characteristics_{jk} + u_{ok} + u_{0jk} + e_{ijk}$$

4. Step: School predictors

Equation 8 Model 4 RQ1: School predictors

$$TSE_{ijk} = \beta_{0ijk} + \beta_{BLOCK1} Teacher\ characteristics_{ijk} + \beta_{BLOCK2} Classroom\ characteristics_{ijk} + \beta_{BLOCK3} Principal\ characteristics_{jk} + \beta_{BLOCK4} School\ characteristics_{jk} + u_{ok} + u_{0jk} + e_{ijk}$$

with:

i = teacher

j = school

k = country

β = predictor

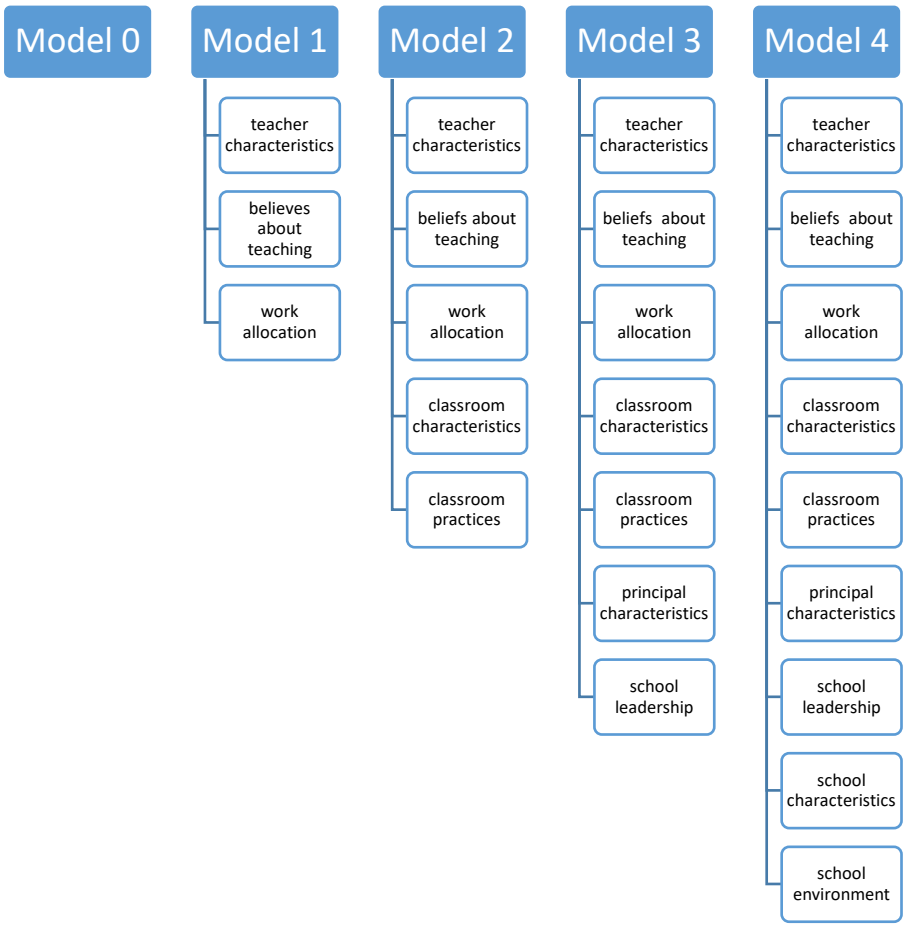
u_{0k} = random term on the country level

u_{0jk} = random term on the school level

e_{ijk} = random term on the teacher level

Figure 3-1 shows the build-up of the first research question.

Figure 3-1 Build-up of the MLM for the first research question



This procedure was repeated for each three-level model for each domain of TSE: teacher's self-efficacy in classroom management, TSE in student engagement and TSE in instruction.

3.5 DESCRIPTION OF VARIABLES

The variables used in this study are drawn from the teacher- and principle- questionnaires of the TALIS 2013 study. As previously described, the data is multilevel, as teacher are nested in schools, which are nested in countries. Teachers provide information on themselves (i.e. socio-demographics, their own teaching or experience as teachers) and the classroom they are teaching in (i.e. student achievement, classroom climate), whereas principles report on themselves (socio-demographics, school leadership and school climate) and the school environment (i.e. location or school composition). Accordingly, teacher and classroom information both operate on the teacher level, principle and school information on the school level. There are no explanatory variables at the country level.

At first, the variables will be described in ascending order starting with the teacher level. A table with the descriptive statistics for each variable in the analysis can be found in Appendix 2.

3.5.1 TEACHER LEVEL (TEACHER AND CLASSROOM CHARACTERISTICS)

Both, teacher and classroom characteristics stem from the teacher questionnaire. The information relies on the of teachers' reporting on themselves and the class with students on ISCED 2 level they are teaching mostly.

3.5.1.1 Teacher characteristics

The teachers were asked to report on socio-demographic information like gender, coded 0-1 (1 = female), work experience (asking for the actual number of years working as a teacher) and educational level[3] (a four-step scale: 1 = below ISCED level 5, 2 = ISCED level 5B, 3 = ISCED level 5A, 4 = ISCED level 6). Teachers age was not included in the analysis, as it was highly correlated with work experience ($r = 0.85$; $p < .001$), which, according to the literature, seems to provide the more coherent explanation for variation in TSE. Furthermore, the teachers reported on their use of time spent in the classroom (actual number of hours/week spent on planning, administrative tasks and teaching). The teacher's constructivist beliefs about teaching consist of four items and were assessed on a four-point Likert scale (with 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree).

3.5.1.2 Classroom characteristics

Teachers were asked to only report on the classroom on ISCED level 2 they spent most time in and which, therefore, is the most representative class for TALIS. They reported on the following eight classroom characteristics: Classroom size (actual number of students in the classroom) and classroom climate (assessed by a four-item scale using a four step Likert-scale

[3] (UNESCO, 1997)
ISCED 2: secondary education first stage/second stage of basic education
ISCED 4: post-secondary non-tertiary education
ISCED 5B: bachelor or equivalent
ISCED 5A: master or equivalent
ISCED 6: doctoral or equivalent

with 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). Items 1, 3 and 4 were negatively phrased, so they were reverse coded. Teachers were asked if they had to wait for a long time for the students to quieten down, whether the students care about a pleasant classroom climate, whether students interrupt a lot and if there is much disruptive noise in the classroom. Student achievement, however, consists of two items asking if 1 = None, 2 = 1%-10%, 3 = 11%-30%, 4 = 31% - 60% and 5 = More than 60% of the students were low academic achievers or academically gifted students. In the same way, a low socio-economic background and the percentage of students speaking a mother tongue, which is different from the language of instruction, were assessed. The teachers also reported on their classroom activities, reporting on the percentage of their total time (max. 100%) spent on administrative tasks, maintaining discipline and actual teaching and learning in the respective class.

3.5.2 PRINCIPAL LEVEL (PRINCIPAL AND SCHOOL CHARACTERISTICS)

The principal and school characteristics both rely on the reporting of the school principal. The source for these is the principal questionnaire.

3.5.2.1 Principal characteristics

The principals reported on their gender (1 = female, 0 =male), their age (actual number of years), the highest educational level they have completed (a four-step scale: 1 = below ISCED level 5, 2 = ISCED level 5B, 3 = ISCED level 5A, 4 = ISCED level 6) and their work experience as a principal in total (number of years). Furthermore, the leadership style of the principle was assessed by using a construct consisting of three items called 'instructional leadership'. For both scales, the items were assessed using a four step Likert scale with the properties 1 = Never or rarely, 2 = Sometimes, 3 = Often and 4 = Very often for instructional leadership and 1 = Strongly disagree, 2 = Disagree, 3 = Agree and 4 = Strongly agree. Instructional leadership, however, consists of three items: Does the principal support co-operation among teachers to develop new teaching practises, take action for new teachers to improve their teaching skills and make sure that teachers feel responsible for their students' learning outcomes.

3.5.2.2 School characteristics

The second type of information the principles reported on are school characteristics. For this study, the following ones were chosen: School size, school type, community size and student socio-demographic information. Schools with only two teachers responding were deleted from the sample.

School enrolment was assessed by asking for the actual number of students registered with the school, school type by asking whether the school is publicly or privately managed (0 = public, 1 = private) and socio-demographic information of the school population by two items asking if 1 = None, 2 = 1%-10%, 3 = 11%-30%, 4 = 31% - 60% and 5 = More than 60% of the students have a different mother tongue than the language of instruction, respectively come from socioeconomically disadvantaged homes.

3.6 QUALITY OF THE SCALES USED

The following section presents the results of the statistical analysis checking for the quality of the scales used. The quality of the scales was tested assessing whether the measurement invariance assumption holds and using confirmatory factor analysis (CFA) to confirm the respective items form a latent construct. At first, quality of the DV 'teacher self-efficacy' will be reported followed by the results of 'constructivist beliefs about teaching', 'classroom climate' and 'instructional leadership style' of the principal.

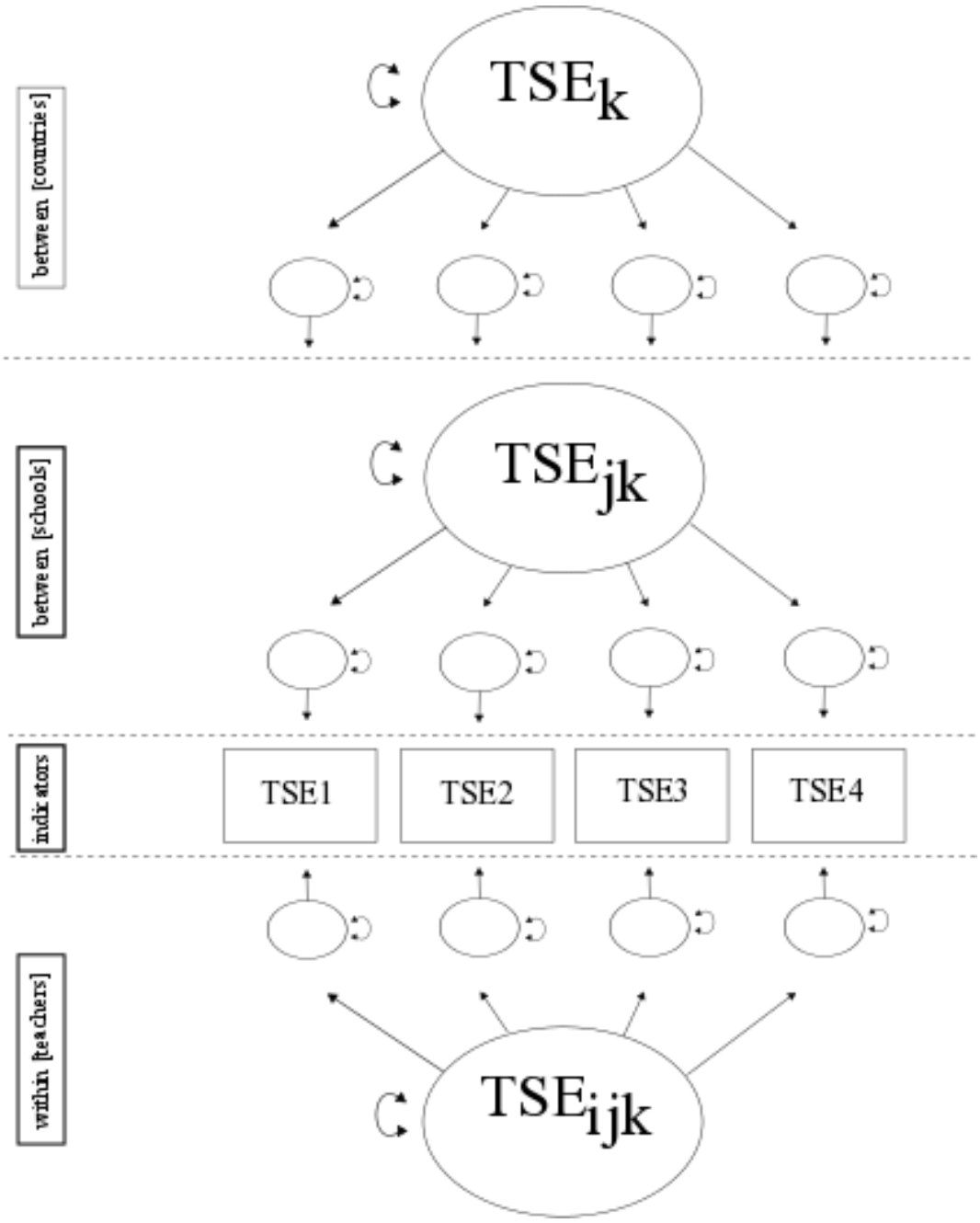
3.6.1 MULTILEVEL STRUCTURE EQUATION MODELLING

With multilevel SEM, the nested data structure of a construct is investigated with respect to the group membership of its level 1 units (Ryu, 2015) and multilevel- and structure equation modelling converge (Kline, 2011). This means that the nesting of the latent variables is followed and thereby both the measurement error of the observed variables and the standard error of the construct are adjusted. If the nested structure of the data is disregarded, the standard errors of the model will be underestimated, and unreliability, which occurs in latent constructs that are measured by several indicators across levels, is overlooked. Furthermore, it enables the researcher to analyse direct and indirect effects (i.e. mediating effects) in one analysis. Another advantage is that on different levels, different

models might occur (i.e. three different types of TSE on level one, but only one factor on level two), which would be disregarded in a single level CFA (Kline, 2011, Muthen, 1991).

The following figure (Fig. 3-2) has been developed to guide my enquiry and shows the basic structure of a multilevel construct using the example of TSE. It is, however, also valid for all other scales assessed using multilevel SEM. The circles represent latent constructs, the rectangles manifest indicators. The paths from constructs to manifest indicators are factor loadings. The smaller circles above and below the manifest variables represent the variance-covariance matrix of within-level centered indicators. The double-headed arrows of the constructs are variances of latent constructs (i.e., TSE) and of smaller latent constructs error variance, and at the school and country levels sampling error (see Marsh et al. 2009). It starts with the lowest level (teachers/within), at which the construct has been measured, going up to the middle level (schools/between 1), where the construct has been aggregated and the highest level (countries/between 2), on which the data has been aggregated by the multilevel formula once more.

Figure 3-2Multilevel structure of teacher self-efficacy



3.6.2 MEASUREMENT INVARIANCE TESTING

Furthermore, invariance models were specified to test whether the scales are invariant across countries and therefore it can be assumed they have the same meaning across teachers, schools and countries in the 3-level CFA; respectively the same construct is measured

across all 32 countries (Millisap, 2011). To test for measurement invariance, a set of CFAs were run, imposing increasingly stricter model constraints on the models, to test whether equivalence holds true, even under strict assumptions (Meredith, 1993). There are three levels of invariance that were established here: First, configural (or pattern) invariance, which does not impose any model constraints on the variables (Meredith & Teresi, 2006). It also serves as the baseline model against which all future models will be compared. The CFI and TLI as well as the RMSEA and SRMR are not supposed to fall below, respectively above, the thresholds for model fit when more constraints are added to the model. These are <.90 for the Comparative Fit Index (CFI) and Tucker-Lewis-Index (TLI), respectively >.08 for the Root Mean Square Error of Approximation (RMSEA) and Standardised Root Mean Square Residual (SRMR). Because of the large sample size and the sensitivity towards it of the χ^2-test, this goodness of fit index will not be used, but the above-mentioned ones (Cheung & Rensvold, 2002) will be used Brown (2006) suggests that a change of < 0.02 in each of the goodness of fit indices show a non-significant change in the quality of the construct and invariance can still be assumed, whereas a change of ΔRMSEA and ΔSRMR < .03, ΔCFI & ΔTLI < .01 (e.g., Cheung & Rensvold, 2002; Chen, 2007) is also deemed acceptable.

Second, metric (or weak) invariance was tested for. Hereby, factor loadings were fixed to be equal across countries. This shows whether differences in the response patterns of respondents are due to differences in the underlying construct. If the construct measures the same concept across countries, the relationships between the items and the construct of respondents within countries are the same. This enables the use of constructs in fixed effects models (i.e., in which the structural part of the model is of interest).

The third level of invariance is called scalar or strong invariance. Scalar invariance additionally requires the relationships between the items and the underlying concepts to be the same for respondents with identical responses across countries. Thus, the intercept of the constructs must be equal across units of analysis in addition to the fixed factor loadings (Meredith & Teresi, 2006). At this level of invariance, the relationship between each unit of analysis (here: teachers) and the DV (here: TSE) are identical. This level of invariance has not been established, as it is not of interest for this study, where fixed effects are investigated, for which the difference of the intercepts is of interest. Scalar invariance would

have, for example, been of interest to investigate a 32-group model, for which I would have had to compare manifest means. Instead, I compare latent means, for which metric invariance is appropriate. The advantage of latent means is that they account for the measurement error associated with the construct and thus provide a more accurate estimate (Millsap, 2011).

As will be seen in the following sections, the model fits for configural variance were sufficient, as well as the Δ _CFI, TLI, RMSEA and SRMR for metric invariance. Consequently, this meets the purpose of the dissertation as I am interested in fixed effects (i.e., the structural part of the models).

It should be noted that this is the first part of a structure equation model (SEM), the measurement model, which asses the mean structure of a construct by linking variables to it (Iacobucci, 2009; Kline, 2010). Model fit is tested by comparing the estimated against the observed variance-covariance matrix, by calculating the (maximum) likelihood. The likelihood has a chi-squared distribution and used in the calculation of goodness of fit indices. To test whether this assumption holds true or not, the difference between these variables and the expected mean is tested. If the suggested construct yields sufficient results, it is shown by fit indices. These should be > 0.95 - > 0.90 for the CFI, < 0.05 - < 0.08 for the RMSEA, > 0.95 - > 0.90 for the TLI and < 0.05 - < 0.08 for the SRMR for the within ($SRMR_W$) and between ($SRMR_B$) levels respectively (Brown, 2015; Tabachnick-Fidell, 2014). The second part of the SEM then is the 'structural model' or 'path model', which links endogeneous (dependent) to exogeneous (independent) variables to test the relationship between these (Iacobucci, 2009).

3.6.3 MEASUREMENT OF THE DEPENDENT VARIABLE: TEACHER SELF-EFFICACY

The measurement used in the present TALIS study is the result of a comprehensive review of the strengths and weaknesses of some of the measurements described above and has been introduced by Tschannen-Moran & Hoy in 2001. It consists of three sub-scales (Efficacy in classroom management, Efficacy in student engagement, Efficacy in instructional strategies), each represented by four items. The teachers were asked to assess themselves

with a four-step likert scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree)[4]:

In your teaching, to what extent can you do the following?

- Efficacy in classroom management
    - [TT2G34D] Control disruptive behaviour in the classroom
    - [TT2G34F] Make my expectations about student behaviour clear
    - [TT2G34H] Get students to follow classroom rules
    - [TT2G34I] Calm a student who is disruptive or noisy
- Efficacy in instruction
    - [TT2G34C] Craft good questions for my students
    - [TT2G34J] Use a variety of assessment strategies
    - [TT2G34K] Provide an alternative explanation, for example, when students are confused
    - [TT2G34L] Implement alternative instructional strategies in my classroom
- Efficacy in student engagement
    - [TT2G34A] Get students to believe they can do well in school work
    - [TT2G34B] Help my students value learning
    - [TT2G34E] Motivate students who show low interest in school work
    - [TT2G34G] Help students think critically

In the first instance, I carried out a three factor CFA (Kline, 2011) that supported the proposed three-factor structure of the measurement for tripartite conceptualisation of TSE (Tschannan-Moran & Hoy, 2001). Then a single level confirmatory factor analysis (CFA) of the a priori three-factor solution within each country as carried out by the TALIS researchers is presented. As the proposed study aims at taking into account the nested structure of

[4] It is probably worthwhile mentioning that the wording has slightly been changed compared to the original items: While the original items ask to what extent teachers agree to the following statements about their teaching, the present items ask to what extent teachers believe they can do some proposed actions. Thus, the wording seems to have been shifted from having confidence in their own teaching to believing in their ability as teachers.

the data and thus a singly level CFA is insufficient, I additionally carried out (1) a two-level CFA for each country and (2) a three level CFA to test for the measurement invariance across countries.

Here, the factor structure of each sub scale was tested in order to check for the use of each sub scale as a single construct as compared to an overall measure or TSE. As the values show, this use is supported:

Table 2 Model fit for all three sub scales of TSE

	CFI	TLI	RMSEA	SRMR within	SRMR b_school	SRMR b_ctry
Student Engagement	.983	.967	.046	.027	.025	.017
CR Management	.992	.985	.035	.015	.021	.078
Instructional SE	.995	.990	.023	.012	.013	.023
Overall	.942	.031	.040	.047	.062	.067

Table 3 below shows the results of the factor loadings for the each of the teacher self-efficacy measures: Self-efficacy in classroom management, in instruction and in student engagement (as done by the OECD TALIS team).

Table 3 Factor loadings for CFA of all countries on a three-factor TSE scale (OECD, 2014) single-level

Scale	*Item*	*Loading*	*Intercept*
Classroom Manage-ment	TT2G34D	1.000	3.279
	TT2G34F	0.080	3.388
	TT2G34H	0.997	3.316
	TT2G34I	1.034	3.227
Instruction	TT2G34C	1.000	3.227
	TT2G34J	1.182	3.133
	TT2G34K	1.065	3.394
	TT2G34L	1.237	3.072
Student Engagement	TT2G34A	1.000	3.218
	TT2G34B	1.098	3.134
	TT2G34E	1.040	2.930
	TT2G34G	0.942	3.091

The following table shows the expansion of the structural validity in a two-level model for each country separately (Table 4).

As the model-fit indices by country show, the factor proposed fitted the structure of the data and confirmed the cross-cultural validity of the instrument (Kline, 2011). This also makes sure that countries can validly be compared to each other in my research, as instruments need to have the same meaning in each country in order to be meaningfully compared (van de Vijver & Leung, 1997). The SRMR is given separately for each level.

Table 4 TSE: Two-level CFA (teachers nested in schools) carried out in each country separately (model fit indices for each country)

Country	CFI	TLI	RMSEA	SRMR_t[5]	SRMR_s	SRMR_c
Australia	.937	.922	.058	.056	.136	.067
Brazil	.948	.938	.045	.045	.086	.067
Bulgaria	.920	.901	.052	.044	.122	.067
Chile	.923	.910	.590	.049	.084	.067
Croatia	.934	.922	.050	.051	.163	.067
Czech Republic	.938	.926	.049	.049	.167	.067
Denmark	.930	.913	.052	.047	.425	.067
Estonia	.910	.893	.055	.054	.169	.067
Finland	.919	.903	.063	.058	.172	.067
France	.923	.908	.050	.050	.181	.067
Israel	.928	.912	.062	.060	.299	.067
Italy	.945	.935	.045	.047	.164	.067
Japan	.955	.946	.046	.052	.089	.067
Korea	.913	.897	.078	.055	.092	.067
Malaysia	.912	.895	.074	.064	.125	.067
Mexico	.964	.957	.035	.037	.129	.067
Netherlands	.951	.941	.041	.039	.206	.067
Norway	.940	.928	.047	.050	.111	.067
Poland	.921	.906	.056	.054	.121	.067
Portugal	.926	.912	.054	.056	.215	.067
Romania	.926	.913	.051	.056	.099	.067
Serbia	.926	.913	.051	.050	.199	.067
Slovak Republic	.916	.900	.060	.051	.134	.067
Spain	.925	.911	.057	.060	.212	.067

[5] _t denotes the teacher level, _s the school level and _c the country level

Sweden	.931	.918	.052	.060	.168	.067
United States	.922	.907	.063	.068	.171	.067
Sub-national entities						
Abu Dhabi (United Arab Emirates)	.934	.922	.053	.049	.178	.067
Alberta (Canada)	.924	.910	.061	.061	.150	.067
England (United Kingdom)	.924	.909	.060	.055	.150	.067

Table 5 shows that most countries fulfil the requirements for the model fit or come close to it; however, between schools the model fit is worse than within schools for most countries (i.e., the $SRMR_{_S}$ is above its proposed threshold of 0.08). For most countries, a three-factor structure on both the teacher and school level, was found to be most appropriate. However, strong correlations on the school level in Australia and Chile suggested only a one factor structure of TSE in these two countries. Moreover, for Bulgaria, Denmark and Israel instructional SE and SE for student engagement could be joined to form only one scale due to strong correlations, too.

The findings show that the 3-factor structure of TSE fits well at the teacher level consistent with the literature (i.e. Tschannan-Moran & Hoy, 2001). However, the goodness of fit at the school level was poorer with SRMRs (0.084 to 0.299) indicating less good fit as a school level construct (e.g., collective TSE). However, as the intention of the dissertation is to carry out both school and country comparisons the country level needs to be taken into consideration. In the next step a three-level CFA was specified (teachers nested in schools nested in countries). As shown in Table 5, the overall model fit for the three-level model was good. This supports the use of the TSE at the teacher, school and country level, as long as all three levels are modelled at the same time.

The following table shows the results for assessing measurement invariance of the TSE scale across the 32 countries in the data set:

Table 5 Model fit for measurement invariance of TSE across 32 countries (c.f. OECD, 2013)

Invariance Level	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SEMR
Configural	.930	.909	.061	.048	---	---	---	---
Metric	.923	.915	.059	.063	.007	.006	.002	.015

The model fits for configural variance were sufficient, as well as the ∆ CFI, TLI, RMSEA and SRMR for metric invariance, warranting the same metric for regression coefficients across levels and countries. The ∆–values reveal metric invariance was the highest level of invariance that was reached by testing the models on the data. Consequently, this meets the purpose of the dissertation as I am interested in fixed effects (i.e., the structural part of the models).

An important feature in using the TALIS measures is the attention given to ensuring validity and reliability of the measurements used to operationalise the construct of TSE and its sub dimensions. My secondary analyses reveal that the internal consistency of all three subscales was acceptable to very good for all participating countries ranging from 0.732 in Bulgaria to 0.872 in Korea for SE in Classroom management, 0.690 in Flanders to 0.847 in Korea for SE in Instruction and 0.726 in Bulgaria and 0.863 in Singapore for SE in Student engagement as shown in Table 6.

Table 6 Reliability statistics for each sub scale of TSE by country

	Classroom Management			*Instruction*			*Student Engagement*		
Country	α	M	(SD)	α	M	(SD)	α	M	(SD)
Australia	.84	3.34	.55	.79	3.31	.53	.84	3.12	.59
Brazil	.80	3.36	.50	.77	3.36	.47	.78	3.40	.48
Bulgaria	.73	3.40	.46	.76	3.19	.51	.73	3.14	.47
Chile	.82	3.40	.52	.80	3.35	.50	.77	3.30	.53
Croatia	.84	3.20	.55	.74	3.26	.50	.74	2.74	.48
Czech Republic	.83	3.01	.58	.72	2.89	.51	.77	2.52	.51
Denmark	.80	3.51	.45	.72	3.35	.46	.76	3.37	.45
Estonia	.79	3.10	.55	.74	2.91	.53	.72	3.04	.50
Finland	.84	3.22	.56	.77	3.03	.56	.82	3.01	.59
France	.80	3.48	.45	.63	3.30	.44	.76	3.14	.48
Israel	.85	3.30	.56	.79	3.25	.55	.81	3.22	.56
Italy	.78	3.39	.48	.74	3.35	.44	.76	3.35	.45
Japan	.86	2.59	.58	.81	2.45	.52	.73	2.14	.42
Korea	.87	2.93	.57	.85	2.89	.55	.84	2.86	.56
Latvia	.80	3.23	.51	.71	3.12	.46	.73	3.02	.48
Malaysia	.84	3.51	.47	.84	3.34	.51	.85	3.44	.48
Mexico	.78	3.24	.55	.77	3.26	.51	.70	3.29	.52
Netherlands	.85	3.28	.52	.64	3.03	.48	.74	2.96	.49

Norway	.83	3.17	.52	.73	2.98	.48	.73	2.75	.46
Poland	.82	3.38	.52	.75	3.07	.51	.77	2.93	.54
Portugal	.81	3.47	.45	.75	3.60	.39	.76	3.52	.42
Romania	.78	3.67	.40	.72	3.67	.37	.76	3.40	.46
Serbia	.80	3.29	.53	.73	3.19	.48	.78	3.02	.53
Singapore	.87	3.19	.61	.84	3.08	.58	.86	3.10	.62
Slovak Republic	.81	3.40	.48	.77	3.24	.47	.78	3.23	.49
Spain	.82	3.13	.56	.75	3.24	.48	.80	2.90	.57
Sweden	.83	3.25	.56	.73	3.17	.51	.75	3.04	.51
United States	.84	3.38	.55	.79	3.33	.54	.85	3.11	.63
Sub-national entities									
Abu Dhabi (UAE)	.80	3.68	.44	.78	3.64	.45	.78	3.62	.46
Alberta (Canada)	.84	3.40	.54	.79	3.30	.53	.84	3.10	.59
England (UK)	.84	3.47	.53	.76	3.41	.49	.82	3.26	.56
Flanders (Belgium)	.85	3.49	.49	.69	3.24	.47	.78	3.14	.51
∑	.82	3.30	.56	.75	3.22	.54	.78	3.12	.59

3.6.4 INSTRUCTIONAL STYLE OF LEADERSHIP

The instructional style of leadership of the principal was assessed by a three-item scale (c.f. here and in the following: OECD, 2013; p. 181ff):

> *Please indicate, how frequently you engaged in the following in this school during the last 12 months:*

- [TG2G21C] I took actions to support co-operation among teachers to develop new teaching
- [TG2G21D] I took actions to ensure that teachers take responsibility for improving their teaching
- [TG2G21E] I took actions to ensure that teachers feel responsible for their students learning outcomes

The unconstrained model fit by country yielded sufficient results. The CFI was different at the country level. It varied between 1.000 (in Australia, Czech Republic, Estonia, Finland, Iceland, Mexico, Netherlands, Singapore, Spain and Abu Dhabi) and .746 (in Bulgaria). The TLI also varied between 1.068 (England) and .238 (Bulgaria), the RMSEA between 286 (Sweden) and .000 (in Australia, Czech Republic, Estonia, Finland, Iceland,

Mexico, Netherlands, Singapore, Spain and Abu Dhabi) and the SRMR between .068 (in Poland, Portugal) and .000 (Spain). Apparently, the scale fits in some countries better than in others, but it can nonetheless be assumed that sufficient fit for comparative analysis was yielded for all of them.

Table 7 CFA for Instructional Style of Leadership by country (c.f. OECD 2013)

	Instructional Leadership style: model fit for CFA by country			
Country	CFI	TLI	RMSEA	SRMR
Australia	1.000	1.043	.000	.003
Brazil	.995	.985	.033	.020
Bulgaria	.746	.238	.210	.059
Chile	.971	.238	.210	.059
Croatia	.999	.997	.023	.022
Czech Republic	1.000	1.004	.000	.017
Denmark	.960	.879	.156	.036
Estonia	1.000	1.004	.000	.015
Finland	1.000	1.033	.000	.011
France	.876	.628	.208	.067
Israel	.970	.909	.110	.040
Italy	.993	.979	.048	.029
Japan	.988	.963	.072	.032
Korea	1.000	1.052	.000	.005
Latvia	.997	.991	.040	.020
Malaysia	.956	.867	.168	.025
Mexico	1.000	1.044	.000	.009
Netherlands	1.000	1.020	.000	.018
Norway	.983	.948	.192	.040
Poland	.878	.633	.191	.040
Portugal	.880	.640	.237	.068
Romania	.975	.924	.115	.034
Serbia	.989	.966	.063	.022
Singapore	1.000	1.007	.000	.015
Slovak Republic	.973	.919	.096	.029
Spain	1.000	1.054	.000	.000
Sweden	.845	.536	.286	.067
United States	1.000	1.000	.000	.000
Sub-national entities				
Abu Dhabi (United Arab Emirates)	1.000	1.024	.000	.004
Alberta (Canada)	.989	.967	.082	.031
England (United Kingdom)	1.000	1.068	.000	.004
Flanders (Belgium)	.967	.902	.181	.032

The reliability statistics for each country, however, show good internal consistency varying between .86 (Abu Dhabi) and .62 (Bulgaria).

Table 8 Reliability statistics for Instructional Style of Leadership (c.f. OECD, 2013)

	Instructional leadership style		
Country	α	M	(SD)
Australia	.83	2.89	.54
Brazil	.78	3.01	.58
Bulgaria	.62	3.04	.42
Chile	.77	3.21	.55
Croatia	.71	2.71	.49
Czech Republic	.77	2.82	.56
Denmark	.72	2.51	.55
Estonia	.70	2.50	.56
Finland	.69	2.45	.52
France	.70	2.64	.55
Israel	.81	2.75	.52
Italy	.72	2.34	.47
Japan	.76	2.99	.55
Korea	.73	2.85	.46
Latvia	.68	3.39	.48
Malaysia	.86	3.17	.59
Mexico	.76	2.70	.57
Netherlands	.71	2.61	.55
Norway	.83	2.99	.47
Poland	.67	2.79	.59
Portugal	.79	3.10	.54
Romania	.76	3.02	.51
Serbia	.69	3.08	.55
Singapore	.79	3.01	.49
Slovak Republic	.66	2.74	.65
Spain	.74	2.56	.61
Sweden	.78	3.15	.66
United States	.88	3.15	.99
Sub-national entities			
Abu Dhabi (United Arab Emirates)	.87	3.37	.58
Alberta (Canada)	.85	3.07	.61
England (United Kingdom)	.78	3.30	.56
Flanders (Belgium)	.83	2.47	.54
Ø	.76	2.89	.56

In contrast to TSE, the construct of instructional style of leadership did not prove to have the same metric invariance and results differed slightly across countries. Configural invariance was the highest level of invariance that could be established as the change of the model fit indices for metric was >.02. The OECD (2013), however, suggests the scale is still suitable for international comparisons, but as the scale might have slightly different meanings in different countries, the results should be interpreted with caution.

Table 9 Measurement invariance results for Instructional Leadership (c.f. OECD, 2013)

Invariance Level	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SEMR
Configural	.980	.939	.097	.034				
Metric	.952	.951	.087	.167	.280	.012	.010	.133

3.6.5 CLASSROOM CLIMATE

Classroom climate was assessed by a four-item scale in TALIS (c.f. here and in the following: OECD, 2013; p. 230ff):

> *How strongly do you agree or disagree with the following statements about your class?*

- [TT2G41A] When the lesson begins, I have to wait quite a long time for students to quieten down.
- [TT2G41B] Students in this class take care to create a pleasant learning atmosphere.
- [TT2G41C] I lose quite a lot of time because of students interrupting lessons.
- [TT2G41D] There is much disruptive noise in this classroom.

The CFI was between 1.000 (in Bulgaria, Czech Republic, Finland, Israel, Japan, Malaysia, Norway, Poland, Romania, Singapore, Spain, USA, Abu Dhabi and Flanders) and .986 (in Chile). The TLI was between.981 (Spain) and 1.007 (Abu Dhabi), the RMSEA between .000 (Croatia, Finland, Japan, Norway, Poland, Romania, Slovak Republic, USA and Abu Dhabi) and .065 (in Croatia) and the SRMR between .001 (in Bulgaria, Portugal) and .022 (Chile). Accordingly, sufficient model fit within countries can be assumed.

Table 10 CFA for Classroom Climate by country (c.f. OECD, 2013)

	Classroom Climate: model fit for CFA by country			
Country	CFI	TLI	RMSEA	SRMR
Australia	.999	.998	.012	.007
Brazil	.999	.998	.012	.005
Bulgaria	1.000	1.002	.000	.001
Chile	.986	.959	.065	.022
Croatia	1.000	1.001	.000	.003
Czech Republic	1.000	.999	.013	.006
Denmark	.995	.985	.041	.012
Estonia	.998	.995	.024	.006
Finland	1.000	1.001	.000	.002
France	.999	.997	.023	.005
Israel	.999	.998	.020	.006
Italy	1.000	1.000	.000	.005
Japan	.999	.997	.032	.006
Korea	.998	.993	.035	.009
Latvia	1.000	1.004	.000	.003
Malaysia	.998	.993	.040	.008
Mexico	.995	.984	.034	.014
Netherlands	1.000	1.003	.000	.004
Norway	1.000	1.004	.000	.002
Poland	.998	.993	.029	.008
Portugal	1.000	1.002	.000	.001
Romania	.995	.986	.045	.011
Serbia	1.000	1.001	.000	.001
Singapore	.994	.981	.054	.011
Slovak Republic	1.000	.999	.011	.005
Spain	.999	.998	.018	.005
Sweden	1.000	1.002	.000	.002
United States	.997	.992	.025	.011
Sub-national entities				
Abu Dhabi (UAE)	1.000	1.007	.000	.003
Alberta (Canada)	.998	.993	.035	.011
England (UK)	.996	.989	.049	.011
Flanders (Belgium)	1.000	1.002	.000	.002

Reliability statistics revealed high internal consistency within countries and varied between .769 (Netherlands) and .882 (England) (Tab. 11).

Table 11 Reliability and descriptive statistics for Classroom Climate

	Classroom Climate		
Country	α	M	(SD)
Australia	.88	2.10	.69
Brazil	.86	2.57	.68
Bulgaria	.85	2.04	.60
Chile	.79	2.41	.64
Croatia	.87	1.97	.60
Czech Republic	.86	2.06	.61
Denmark	.82	1.96	.62
Estonia	.88	2.06	.65
Finland	.88	2.22	.68
France	.86	2.26	.70
Israel	.86	2.03	.57
Italy	.84	1.79	.59
Japan	.85	2.19	.55
Korea	.81	2.17	.60
Latvia	.83	2.16	.62
Malaysia	.87	2.04	.59
Mexico	.77	2.30	.57
Netherlands	.83	2.61	.55
Norway	.82	1.99	.60
Poland	.85	2.31	.67
Portugal	.88	1.90	.56
Romania	.82	1.98	.62
Serbia	.86	2.35	.65
Singapore	.87	2.24	.61
Slovak Republic	.86	2.40	.70
Spain	.87	2.17	.71
Sweden	.86	2.13	.66
United States	.65	3.19	.69
Sub-national entities			
Abu Dhabi (UAE)	.81	1.93	.60
Alberta (Canada)	.85	2.12	.62
England (UK)	.88	2.04	.70
Flanders (Belgium)	.86	2.25	.67
Ø	.84	2.18	.63

As the change parameters for measurement invariance show, it was possible to establish configural and metric invariance for the perceived classroom climate.

Table 12 Measurement invariance for Classroom Climate across 32 countries (c.f. OECD 2013)

Invariance Level	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SEMR
Configural	0.999	0.997	0.02	0.007				
Metric	0.991	0.989	0.04	0.06	0.008	0.008	0.02	0.053

3.6.6 CONSTRUCTIVIST BELIEFS ABOUT TEACHING

Teacher's constructivist beliefs about teaching were assessed by a four-item scale in TALIS (c.f. here and in the following: OECD, 2013; p. 237ff):

> *We would like to ask about your personal beliefs on teaching and learning. Please indicate how strongly you agree or disagree on the following statements?*

- [TT2G32A] My role as a teacher is to facilitate students' own inquiry.
- [TT2G32B] Students learn best by finding solutions to problems on their own.
- [TT2G32C] Students should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved.
- [TT2G32D] Thinking and reasoning processes are more important than specific curriculum content.

The CFI was between .916 (USA) and .999 (in Czech Republic, Estonia, Mexico, Norway). The TLI was between.747 (USA) and .998 (Estonia), the RMSEA between .014 (Czech Republic) and .122 (Israel) and the SRMR between .006 (Singapore) and .031 (Japan). Although not all model fit parameters revealed optimal fit, this scale was deemed to show sufficient quality for international use.

Table 13 Model fit for CFA by country for Constructivist beliefs about teaching (c.f. OECD 2013)

	Constructivist Beliefs about teaching: model fit for CFA by country			
Country	CFI	TLI	RMSEA	SRMR
Australia	.998	.994	.018	.009
Brazil	.957	.871	.054	.029
Bulgaria	.992	.976	.030	.015
Chile	.974	.923	.083	.022
Croatia	.981	.942	.067	.020
Czech Republic	.999	.997	.014	.007

Denmark	.998	.994	.018	.009
Estonia	.999	.998	.010	.007
Finland	.985	.954	.056	.018
France	.983	.948	.060	.019
Israel	.953	.859	.089	.031
Italy	.974	.923	.073	.024
Japan	.996	.987	.040	.010
Korea	.974	.922	.074	.027
Malaysia	.999	.997	.014	.009
Mexico	.995	.986	.035	.010
Netherlands	.992	.975	.022	.013
Norway	.984	.952	.056	.018
Poland	.981	.943	.063	.018
Portugal	.992	.975	.039	.012
Romania	.998	.995	.017	.006
Serbia	.989	.966	.064	.016
Slovak Republic	.990	.970	.047	.015
Spain	.975	.924	.080	.022
Sweden	.916	.747	.095	.030
United States	.998	.994	.031	.005
Sub-national entities				
Abu Dhabi (United Arab Emirates)	.983	.950	.057	.020
Alberta (Canada)	.986	.957	.056	.018
England (United Kingdom)	.998	.995	.021	.008
Flanders (Belgium)	.998	.995	.017	.008

Measurement invariance analysis showed that configural and metric invariance apply to the data.

Table 14 Measurement invariance results for Constructivist Beliefs about Teaching across 32 countries (c.f. OECD, 2013)

Invariance Level	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SEMR
Configural	.982	.947	.058	.019				
Metric	.966	.959	.051	.065	.016	.012	.007	.046

Reliability statistics revealed high to moderate internal consistency within countries and varied between .541 (Portugal) and .843 (Latvia) (Tab. 15).

Table 15 Reliability and descriptive statistics for Constructivist Beliefs about Teaching

	Constructivist believes about teaching		
Country	α	M	(SD)
Australia	.71	3.10	.48
Brazil	.69	3.09	.53
Bulgaria	.62	3.34	.44
Chile	.79	3.31	.55
Croatia	.73	3.23	.45
Czech Republic	.72	3.16	.43
Denmark	.67	3.28	.45
Estonia	.65	3.13	.41
Finland	.66	3.18	.41
France	.69	3.19	.49
Israel	.76	2.98	.49
Italy	.66	3.12	.41
Japan	.69	3.35	.50
Korea	.84	3.27	.50
Latvia	.71	3.13	.49
Malaysia	.75	3.29	.52
Mexico	.68	3.10	.44
Netherlands	.72	2.98	.39
Norway	.54	3.15	.44
Poland	.71	3.30	.45
Portugal	.71	3.27	.51
Romania	.75	3.27	.47
Serbia	.69	3.28	.46
Singapore	.80	3.21	.45
Slovak Republic	.72	3.13	.51
Spain	.74	2.88	.46
Sweden	.58	3.19	.48
United States	.76	3.19	.41
Sub-national entities			
Abu Dhabi (UAE)	.75	3.41	.48
Alberta (Canada)	.73	3.23	.46
England (UK)	.72	3.19	.46
Flanders (Belgium)	.66	3.18	.43
Ø	.71	3.19	.46

3.6.7 SUMMARY

The analyses that the OECD and I conducted revealed that the scales show sufficient quality for international comparative use. Classroom climate, constructivist beliefs about teaching and teacher self-efficacy came up with promising results for the proposed study. Instructional style of leadership, however, had good measurement invariance results for some countries, but showed rather poorer fit for others. So it has to be acknowledged it may not fit all the countries equally well. But it was possible to yield at least metric invariance for all other scales, which proves that the factor model fits the data simultaneously and thus is a sufficient criterion if latent constructs are to be assessed (Vanderberg & Lance, 2000; Gergorich, 2006), which is the method of analysis used in my research. With respect to the country analysis, it should be noted that previous studies found caveats when comparing Likert scales across countries from different cultural backgrounds (i.e. Heine et al., 2001): The bottom and top response options of Likert scales serve as a point of reference for implicit comparisons with a reference group of presumably similar people and thus do not provide context-free information, as in some countries it is rather uncommon to use the extreme ends of a scale, in others it is not. These differences in response styles according to cultural values like 'individualism' (i.e. North America) and 'collectivism' (i.e. Asian countries) could not solely be attributed to country membership, but also to whether respondents live in their country of origin or if they live abroad. Those living abroad revealed greater within-group differences compared to those living at home, who, in contrast, showed larger between-group differences. Accordingly, differences in responses cannot necessarily be attributed to the cultural origin, but also to environment against which the responses are contrasted (Heine et al., 2002). Another factor to take into account is the behaviour towards the use of the Likert scale: Western countries tend to use the extreme ends of the scale rather than Eastern countries, which tend to use the middle spectrum of response choices (Chen et al., 1995; Lee et al., 2002; Hamamura et al., 2008). Apart from the use of the spectrum of the Likert scale, there is variation in the behaviour towards the questionnaire: Japanese respondents were found more likely to skip questions than i.e. American or Chinese respondents (Lee et al., 2002). That differing response patterns across countries exist and therefore that mean comparisons in this study are not valid has been shown by the measurement invariance testing in this section as it was in most cases only

possible to establish metric invariance. However, as metric invariance was achieved in most cases, this enables the interpretation of beta coefficients across the levels to be based on the same metric.

The internal consistency for all scales was either good (> 0.6) or very good (> 0.8) for all four scales. Accordingly, it can be assumed all scales yield sufficient statistical properties for the proposed study.

3.7 MISSING DATA

When dealing with social science survey data, an issue that frequently occurs are missing values: Study participants do not fill in the whole questionnaire, but skip items, questions, scales or even sections. This causes problems, as it reduces the available information. According to Rubin (1976), there might be a number of different underlying reasons that lead to missingness. These mechanisms each describe a different possibility of data being available or missing: Data missing completely at random (MCAR), where missing data are unrelated to observed data, missing at random (MAR), where missing data depend on observed responses, or missing not at random (MNAR), where missing data depend on unobserved responses (Molenberghs et al., 2015). These three mechanisms, however, each require a different analytical rationale in order to obtain valid results, as the type missingness biases the data distribution (ibid.).

In the following section, I will explain the mechanisms of missingness that can occur in a survey such as TALIS and approaches to deal with them with respect to their strengths and weaknesses. Furthermore, I will provide a description of my approach to missing data in secondary analysis of the TALIS data set.

3.7.1 MISSING DATA MECHANISMS

As described above, there are three different types of missing data mechanisms. In the following, these will be described in more detail and their implications for statistical analysis and power discussed.

3.7.2 MISSING COMPLETELY AT RANDOM (MCAR)

Firstly, data can be missing complete at random (MCAR). MCAR is based on the assumption that the missing values are not dependent on any other value of the data set. In other words, if data are MCAR, all remaining data can be assumed a random sample of the complete data. Moreover, if the data has been as a random sample from the target population, the complete cases can also be seen as a random sample from the population. Accordingly, the available cases can be analysed with no further treatment of missing values and still yield reliable estimates (Molenberghs et al., 2014). For example, if 'age' is missing for some respondents in the study, it is missing MCAR if it cannot systematically be related to any other variable in the data set, i.e. gender or nationality. Formally, this is stated as:

Equation 9 Data missing completely at random

$$P(R|Y_o, Y_m, X) = P(R)$$

(c.f. Mohlenberghs et al., 2014)

with:
P = Probability
R = missing value
X = covariate
Y_o = observed value
Y_m = missing value

If, nonetheless, a full dataset is required, data missing MCAR can easily be imputed using a range of methods (Enders, 2010) (See section 2.11 for a description of imputation methods).

3.7.3 MISSING AT RANDOM (MAR)

Weaker assumptions on the independence of missing and observed values are required by the concept of 'missing at random' (MAR). If there are, however, dependencies among missing and non-missing values, which are based on observed values and the probability

of data missing is equally based on only one out of several other variables, the data points are missing at random (Molenberghs et al., 2014). First, this does affect the distribution of the respective variables in a way that they cannot be assumed equal any more to the distribution in the target population. Second, the remaining complete data represents a biased sample from the target population, so analysing only complete cases would lead to biased estimates. But, however, as the reason for missingness is known, it is possible to generate reliable estimates for the missing values (Enders, 2010). In this example, the missing values for 'age' could be dependent on gender (i.e. women skip the age question more often than men), but by controlling for the effect of a third variable, i.e. educational background, this can be accounted for. Formally, data MAR is noted as:

Equation 10 Data missing at random

$$P(R|Y_o, Y_m, X) = P(R|Y_o, X)$$

(Allison, 2011)

This mechanism of missingness is also called 'ignorable', as it does not require to model the dependencies of missingess, because this information is provided by the present data (Allison, 2012). This is not to be confused with the missingness as such being ignorable; the missing values still need to be replaced with reliable estimates in order to gain reliable results, as the remaining ones – in contrast to MCAR – cannot be seen a random sample of the complete dataset and therefore contain bias that needs to be accounted for (ibid.).

3.7.4 MISSING NOT AT RANDOM (MNAR)

The third mechanism according to Rubin (1976) is 'missing not at random' (MNAR). In this case, the missingness is not random in a probabilistic sense, thus not dependent on available data, but on data that has not been observed (Enders, 2010). This might be, for example, that people do not answer questions about their body weight not based on their gender, which has been observed, but on height, which has not been collected in the questionnaire. In other words, people with a higher BMI technically answer the body weight question less often than people with a lower BMI, but this information remains hidden as

the height has not been assessed. Formally, the relationship between missing and observed values is described as:

Equation 11 Data missing not at random

$$P(R|Y_o, Y_m, X) = P(R|Y_m, X)$$

(c.f. Mohlenberghs et al., 2015)

In contrast to data MAR, if the data are MNAR, the cause of missingness when generating estimates is non-ignorable, as the bias inferred does not allow for generalising from the sample to the whole population (Molenberghs et al., 2014). Accordingly, the missing data mechanism has to be modelled, which requires solid knowledge of the data (Allison, 2002).

3.8 APPROACHES TO HANDLING MISSING DATA

With the increasing use of large-scale data sets over the past 6 decades, various ways of treating missing values have been developed. They can broadly be grouped into two categories: The first (and older) category is called 'complete and available case analysis', which means that only those cases that do not have missing values are used for analysis, and cases with missing values are deleted, respectively disregarded. This approach, however, has certain limitations and is therefore only recommended in the case of very little missingness. The second category, in contrast, comprises different imputation methods that aim at providing unbiased and efficient estimates, (Graham et al., 2021) where a value is missing that are close to the true value and show little variability (ibid.). In the following, I will briefly describe and discuss methods of both approaches.

3.8.1 DELETION METHODS: COMPLETE AND AVAILABLE CASE ANALYSIS

Available case analysis draws on only those participants that provide a full record of answers for the variables of interest and all other participants are disregarded. Moreover, this is done in two different ways: Either list wise or pairwise deletion.

3.8.1.1 List wise deletion

List wise deletion only uses those cases that provide a full record of information on all variables of interest and excludes all other participants from the analysis. From this approach, a few problems occur: first, especially if the data set contains many participants, it is very unlikely that there is a full record for many participants and therefore the number of observations in the analysis might shrink quite substantively. Accordingly, the selected sample in the data set might be very biased, as none of the missing data mechanisms described above are taken into account and the reasons for participants not answering the whole questionnaire might be very different (Graham et al., 2012). Graham et al. (2012) also point out that in a multiple regression, there is no bias inferred if the values on the dependent variable are MCAR, and on the independent variables, only available cases are used, as the cause of missingness is accounted for by the complete cases. In more complex analyses, however, this is no longer true and there is still bias. Accordingly, by applying list wise deletion, easily explanatory power is lost. As described above, from that point of view it is only acceptable, if the data are missing MCAR. Second, it realistically happens that – especially in big data sets – missing values occur not only on a few cases, but rather across many different ones. Therefore, when deleting cases list wise, statistical power is lost, as it partly relies on sample size (Schafer&Graham, 2002).

3.8.1.2 Pairwise deletion

Pairwise deletion, in contrast, draws a new sample of complete cases for each variable in the analysis, so there are different sample sizes within an analysis. As list wise deletion, it is deemed acceptable for many linear models like regression or complex structure equation models (Allison, 2011), but only if the data fulfil certain criteria: First, the data has to be missing MCAR if the analyses should yield reliable estimates; at least in large data sets. Second, if the data is MAR but not observed at random[6], the estimates might be severely biased, for the reasons already mentioned above. Compared to list wise deletion, however,

[6] 'Observed at random' means that a random sample is drawn from the population, in contrast to a convenient sample, which is not selected at random. Missing at random refers to the reasons for missingness in the data independent from the sampling mode

it has the advantage of not losing as many cases, but is still likely to over- or underestimate standard errors, as it uses a different sample size for each standard error and will thus not produce a common covariance matrix (Allison 2012, Graham&Schafer, 2002). This matrix often is not positive definite, which makes it impossible to use the remaining data for analyses that are based on regression.

3.8.2 SINGLE IMPUTATION TECHNIQUES

A different logic is applied to imputation techniques: Instead of either deleting incomplete cases or drawing a new complete sample for every variable, a variety of imputation methods draw on the information provided by available data in order to provide reliable and efficient estimates of the missing values. Some researchers, however, are concerned that imputing data is similar to cheating as data would be 'made-up'. But Enders (2010) points out that the aim of any statistical analysis is to make inferences from the sample to the population, for which imputation leads to enhance the quality of the present data. The following chapter will give a brief overview of the relatively commonly used imputation techniques: mean imputation, regression imputation and hot-deck imputation and their strengths and weaknesses. The focus, however, will be on the second part of the chapter, where multiple (MI) and full information likelihood (FIML) imputation will be introduced, as these are relevant to the current study.

3.8.2.1 Mean imputation

With mean imputation, each gap in the data is replaced with the arithmetic mean of the respective variable. Although this technique results in the desired result of a complete data set, it distorts the variation and standard deviation of a variable as well as the covariance matrix, as the missing value is filled in at the centre of the distribution (Baraldi & Enders, 2013; Graham et al. 2012). The advantage of a complete data set is in this case overthrown by the disadvantage of underestimated statistical values.

3.8.2.2 Regression imputation and stochastic regression imputation

Regression imputation draws on present information in the data set from complete cases to produce reliable estimates of the missing value and thus have a complete data set. This

method results in more reliable estimates than mean imputation, but has two main weaknesses: first, it requires one equation per missing data pattern and when the data set and the missing data pattern both are complex, the analysis becomes complicated to implement (Enders & Peugh, 2009). Second, there is a linear relationship between the imputed values and the dependent variable, which means that correlations between these variables are overestimated and the data set lacks variability that would have been there in the case of originally complete data (Baraldi & Enders, 2013). There are, however, correction processes available if the missingness is MCAR or MAR in order to gain consistent estimates of the covariance matrix and thus achieving estimates that are closer to the true value. Accordingly, it is superior to mean imputation but still inferior to more complex imputation methods (Rubin & Little, 2014).

A second type of regression imputation is stochastic regression imputation. In addition to predicting values of incomplete cases by using information from complete cases, stochastic regression adds a normally distributed residual term, whose distribution has a mean of zero and a variance equal to the variance of the dependent variable, to each predicted value so they gain variability (Baraldi & Enders, 2013). This is superior to linear regression imputation, as the imputed scores gain variability and the parameter estimates are even unbiased when the data is only MAR, but still has the disadvantage that with complex missing data patterns requiring their own equation, the procedure is difficult to implement. Furthermore, as with any single imputation method, stochastic regression imputation tends to underestimate standard errors, as the following statistical analysis does not make a difference between original and imputed scores by, for example, correcting for possible inflation. Accordingly, results are prone to leading to type I errors and imputed values can be out of bounds (ibid.).

3.8.2.3 Hot-deck imputation

Hot-deck imputation is comparitively rarely used in the social sciences, but more commonly used in public data sets. As the TALIS data is a large-scale survey that shares characteristics with public data sets, it will, nonetheless, be mentioned here. The basic principle of hot-deck imputation is that it replaces each missing value with either a random draw of

values from other participants with similar response patterns or certain demographic characteristics (random hot deck method) or the values are taken from a respondent, who has been identified as a close neighbour according to pre-defined metrics (deterministic hot deck method) (Andridge & Little, 2010). It maintains the univariate distributions of the data as well as not reducing its variability to the extent other methods do. This imputation method, however, shares the downside of other imputation methods to underestimate standard errors (Baraldi & Enders, 2013). Its particular downside for social science research is that the parameters it produces in regression and correlation measures of association are biased (Graham & Schafer, 2002) and thus it is not suitable for the present study.

3.8.3 FULL INFORMATION MAXIMUM LIKELIHOOD IMPUTATION (FIML) AND MULTIPLE IMPUTATION (MI)

After discussing several comparatively straightforward methods of imputing missing data in the previous sub-sections, which all, due to their weaknesses, have proven not to be suitable for the present study, the following section focuses on the more complex and - in this case - advantageous methods of FIML and MI. In contrast to the previous sub sections, it will provide more in-depth discussion and also explain the underlying estimation and imputation mechanisms.

3.8.3.1 Full information maximum likelihood imputation

If the data are MAR it is possible to use (full information) maximum likelihood information to gain estimates for the missing values. The FIML approach is based on drawing inferences from a likelihood function, given the model for the complete data is realistic (Schafer & Graham, 2002), i.e. to maximize the probability to observe through estimation what would have been observed, if the data were present (Allison, 2002).

Maximum likelihood estimation comprises a variety of analytical methods, which are suitable for analysis as long as the data stems from a multivariate normal distribution (ibid.). On the one hand, in contrast to the imputation methods mentioned above, FIML estimation does not replace the missing values, but rather estimates parameters and adjusts standard errors for the missing values by using the data that are available. Deletion

methods, on the other hand, tend to lead to overestimated parameters, as they disregard possible correlations between the deleted and remaining variables and therefore information is lost. Using ML improves the accuracy of parameter estimates by utilizing information contained in both the observed and unobserved data (Enders, 2010). This is achieved by the Estimation Maximization algorithm (EM), which increases the likelihood of the available data matching the complete data until convergence has reached a local maximum (Rubin & Little, 2002). It basically consists of a chain of iterative steps: (1) Missing values are replaced by estimated values, (2) parameters are estimated, (3) missing values are re-estimated under the assumption that the new parameters are correct, (4) parameters are re-estimated… until the model convergences (ibid.). Hereby, the first step is called expectation, as the estimated value for all variables with missing data are computed, given that there is a true value for both, the population mean and the variance-covariance matrix. In the following maximization step, the expected value is utilised to estimate the population mean and covariance (Pigott, 2001). However, Pigott points out that the lacking ability to calculate standard errors limits the usefulness of the method. Larsen (2011), nonetheless, could show that for multilevel data FIML imputation yields less biased estimates both under MCAR and MAR missing data than multiple imputation for the higher order units. However, despite Larsen's promising results, MI is still deemed more appropriate as Larsen's results are based on longitudinal data, whereas the present data are cross sectional.

3.8.3.2 Multiple imputation

The second imputation method considered to be state-of-the-art is multiple imputation (MI). During MI several data sets containing plausible values for the missing ones are generated. Thus, the procedure results in multiple complete data sets, which can each be analysed using standard statistical methods. But Enders (2010) points out that the aim of any statistical analysis is to make inferences from the sample to the population, for which imputation yields to enhance the quality of the present – or, as Todd Little (2014) roughly puts it: it is worse to not impute the data due to the weaknesses mentioned, than to impute it. Furthermore, when comparing FIML and multiple imputation – mostly in the case of large-scale data – both techniques result in the same (or at least very similar) parameter

estimates, which are not significantly different from each other. Accordingly, MI can be deemed equally trustworthy if applied properly. Finally, MI produces several possible values for each missing, whereby it accounts for the uncertainty that is associated with the missing data. As MI simulates sampling by drawing multiple times from the same distribution, it is possible to account for the sampling error and adjust standard errors when findings from the models run with different imputed datasets are merged together. The main advantage MI has over other imputation techniques such as FIML is that the uncertainty around the missing value is taken into account (ibid.).

As FIML, MI yields reliable estimates if the data are either MCAR or MAR. It is based on a Bayesian process, which consists of three phases. In the first phase, the data are imputed by estimating several plausible values for each missing value. It consists of two steps: In the first step, the imputation step, values are predicted through regression equations by using a covariance matrix (Peugh & Enders, 2009). For each missing data pattern, a set of regression equations is specified and through multiple iterations, estimates are calculated.

Collins et al. (2001) speak of the 'imputation model' in contrast to the 'analyst model' to consider that there are two separate stages. The imputation model refers to the first stage, where data are modelled so that the imputation can be carried out thoroughly. After that, a different model is applied to the data that is of interest to the analysis of the research question; typically, more variables are used in the MI than in the analysis. This comes with both advantages and disadvantages. The main advantage is that it allows the use of auxiliary variables to enhance the accuracy of the imputation, which are not necessarily of interest to the analysis and model those separately. This makes the use of MI very flexible and allows for considering that imputing and analysing the data are two separate procedures. The disadvantage that occurs is that the imputation model can easily be 'over specified' in a way that it is no longer general enough so that the imputation results represent data that do not allow the construction of an analytic model that suggests relationships that are different from each other between the variables of interest. Accordingly, the challenge here is to choose auxiliary variables that add information to the imputation but do not substantially bias the results. Moreover, MI operates within the Bayesian framework

by applying prior parameters to the model against which the posterior distribution is compared. This is to take into account prior/existing knowledge about the model of choice into the compilation of the imputed data.

The 'analyst model', then, is the model which is applied to the data in order to answer the research question (ibid.).

These estimates, however, never converge to a single value but the process is repeated unless the data augmentation converges, and the distributions no longer vary systematically. This is supposed to happen after at least 50 iterations (Enders, 2010). MPlus, however, saves estimates that occur after every 100th iteration (Muthén & Muthén, 2012). In the second step, the estimates are amended by a random draw from the normal distribution of residuals in order to account for residual variation. Thereby, it uses a different set of estimates of the population covariance matrix for each of the data sets. Multiple data sets are produced by comparing the newly calculated values to those from the previously imputed ones. This is based on the Bayesian principle of first defining a prior distribution, then estimating a likelihood function and finally defining the posterior distribution, which is the observed distribution corrected for the prior distribution (Enders, 2010). Graham, Olchowski & Gilreath (2007) demonstrated that usually five data sets yield sufficiently unbiased results; although they recommend using >20, but this often results in an extreme increase in computational effort and computing power. Using fewer data sets, however, could lead to underestimated standard errors. In the second phase the data are analysed using the desired statistical method and in the third phase, the analysed data are combined into a single result (Rubin, 1987; Enders, 2010).

3.8.3.3 MI and Multilevel data

Hence, MI provides the advantage over many of the simpler imputation methods that it is able to account for multilevel structure of data (van Buuren, 2011). If the multilevel structure of the data is ignored, typically standard errors are underestimated, and confidence intervals are too small. This can be adjusted for by either splitting the data file according to its levels and then imputing each level in a separate file to disentangle the multilevel structure first and then merging the files after successful imputation again. However, van

Buuren's approach ignores dependencies between the levels in the data and useful information for the information is lost that way. Thus, the dependency will be used as a tool for analysis and the multilevel structure was explicitly specified as it was done here in my research using MPlus. This allows for imputing all data at once.

3.9 IMPUTATION OF MISSING VALUES

In the whole TALIS data set used in my research, between <.01% (teacher gender) and 19.1% (hours spent on either teacher and learning, keeping order or admin tasks) of the observations are missing (See Table 19). Little's test, which tests the assumption whether the data are MCAR, was carried out for the groups of teacher-, classroom-, principal- and school variables separately. All four tests were highly significant showing that none of the variables are missing MCAR[7]. This is also supported by the patterns of missingness (see Section 3.14ff). According to Little and Rubin (2014) it is impossible to empirically distinguish whether the data are missing MAR or MNAR. It is assumed the data are MAR. Missing values were imputed using multiple imputation in MPLus (Muthén & Muthén, 2012). The programme was set to generate five data sets with, in total, five estimates for each missing value in order to provide accurate replacements. For creating each estimate, it uses 100 iterations. Although it is argued that at least 20 data sets containing plausible values should be produced, as only between <.01% and 19.1% (between 3 and 19,925 observations) of the data are missing, five seem to be sufficient in order not to lose too much statistical power. This is supported by Rubin (1987), whose concept of relative imputation efficiency suggests that for less than 10% missingness 5 data sets provide > .9804 relative efficiency (RE) and for 20% missingness >.9615 relative efficiency. The RE describes the accuracy of the data after imputing a certain number of parameters for each missing value depending on the percentage of missingness:

[7] *Teacher variables*: Chi-square: 23265,320; df = 16659; Sig. > .001
Classroom variables: Chi-square: 7423,450; df = 1612; Sig. > .001
Principal variables: Chi-square: 16320,433; df = 623; Sig. > .001
School variables: Chi-square: 785,600; df = 46; Sig. > .001

Equation 12 Relative efficiency of imputed values

$$RE = {}^{1}\Big/_{1+\frac{\lambda}{m}}$$

with:

λ = % missing values

m = number of imputed data sets

As Table 19 shows, most variables have less than 10% missing values; only five variables on the teacher level (Class size, Classroom climate, hours spent teaching and learning, hours spent on administrative tasks, hours spent keeping order in the classroom) have between 10%-20% missing values and one variable (principal's work experience) has slightly more than 10% values missing. For the five imputed datasets the relative efficiency in my study is .9632, which is sufficient. The minimum threshold of .95 was thereby reached. Table 18 shows the RE for different amounts of missing data and number of imputed data sets:

Table 16 RE for different ms and λs (own table after Horton & Lipsitz, 2001 p. 246)

m	10%	20%	30%	50%	70%
3	.9677	.9375	.9091	.8571	.8108
5	.9804	.9615	.9434	.9091	.8772
10	.9901	.9804	.9709	.9524	.9346
20	.9950	.9901	.9852	.9756	.9662

Furthermore, the time the statistical analysis of the model took set another constraint. Due to limited access to powerful computers, the last model took about 24 hours to converge with only 5 data sets and would have taken more time than available for all three subscales when having more than 5 data sets.

Table 17 Missing values per variable

Teacher	**Valid**	**Missing**	**%**
Female Gender	104,285	3	.003
Educational Level	102,841	1,447	1.407
Teaching Exp.	97,709	6,579	6.733
Hours Teaching	101,433	2,855	2.815
Hours Planning	100,249	4,039	4.029

Hours Admin. Work	94,479	9,809	10.382
Construct. Bel.	99,560	4,728	4.749
Classroom			
Class size	85,606	18,682	21.823
Classroom Climate	85,989	18,299	21.281
Low Student Achievement	99,374	4,914	4.945
Teaching & Learning	84,363	19,925	23.618
Keeping Order	84,363	19,925	23.618
Admin. Tasks	84,363	19,925	23.618
Different Language	99,483	4,805	4.830
Disadv. Background	99,257	5,031	5.069
	104,288		
Principal			
Female Gender	6,090	330	5.419
Age	6,071	349	5.749
Educational Level	6,059	361	5.958
Work Experience	5,704	716	12.553
Instructional Leadership	6,005	415	6.911
School			
Student Enrolment	5,993	427	7.125
Type of School	6,084	336	5.523
Community Size	6,062	358	5.906
	6,420		

The imputation was carried out in MPlus using the Bayesian imputation model described in Section 3.8.3.2. Due to data size limitations of MPlus and the complexity of the model, each country was imputed separately using a two-level model. Although it would have been the favourable approach, a three-level model was not possible due to data size and model complexity restrictions of MPlus. The attempt to impute all values using a three-level model resulted in an empty output after four weeks of calculation. For three variables on the principal level in the Netherlands, Norway, Romania, Slovak Republic, Spain and Sweden, however, a two-level imputation was not possible, as there was no within-country variation, which is a requirement for MPlus to estimate imputed values, which are based on the variance of the respective variable. The variables are namely: Highest level of education (master degrees only), School environment (medium-sized towns only) and School type (public schools only). For these three variables, the MCMC imputation tool in SPSS

was used and a single-level imputation carried out. SPSS thereby starts from the first variable with missing values and calculates in each iteration one possible value for each missing one, till the process results in a complete data set (IBM, 2016).

The data contain three categorical variables, teacher gender, principal gender (male vs female) and type of school (public vs private). As with big data sets and complex models the imputation of categorical variables often fails, the variables weree treated as continuous and then rounded to 0 or 1 using the .5 cut-off (Yucel & Zaslavsky, 2001; Graham, 2009).

The imputation conducted for my research resulted in creating five full data sets each containing information from 104.288 teachers and 6.420 principals. After the imputation had finished, the countries were combined into five data sets each. In total, the 5x32 models took about two weeks to converge.

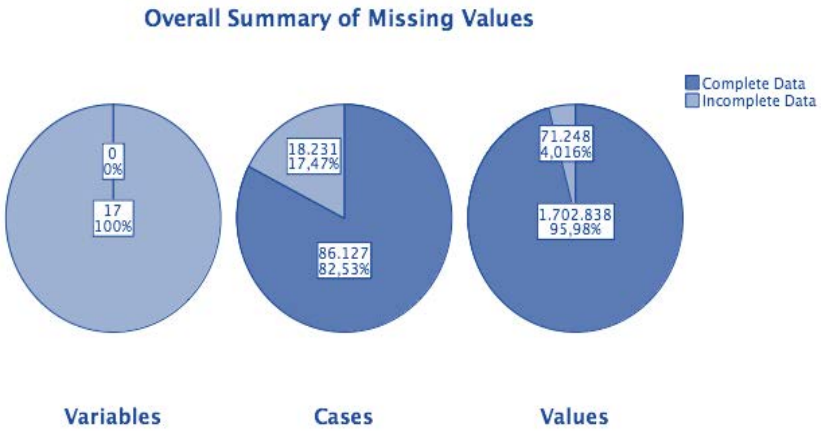
3.10 MISSING VALUE PATTERNS AND STATISTICS

In the following, graphical representations of missing data for variables in my research are presented. The structure is found to be equally valid for teacher, classroom, principal and school characteristics. Firstly, pie charts display missingness by cases, variables and values. Secondly, a summary of missing value patterns shows whether there is a systematic variation in the missingness or not and thirdly, the ten most frequently occurring missing data patterns are shown in the selected bar charts. It is important to bear in mind that the information about the teachers and the classroom both stem from the same source, i.e. the teachers. The information about principals and schools, in the same way, stem from the principal. Accordingly, the separation below into teacher, classroom, principal and school characteristics is merely practical; from a logical perspective, the teacher and classroom as well as the principal and school variables could have been investigated together.

3.10.1 TEACHER VARIABLES

For the teacher related variables, least missingness occurs by variable (< 1%), and out of these 1%, most by cases/participants.

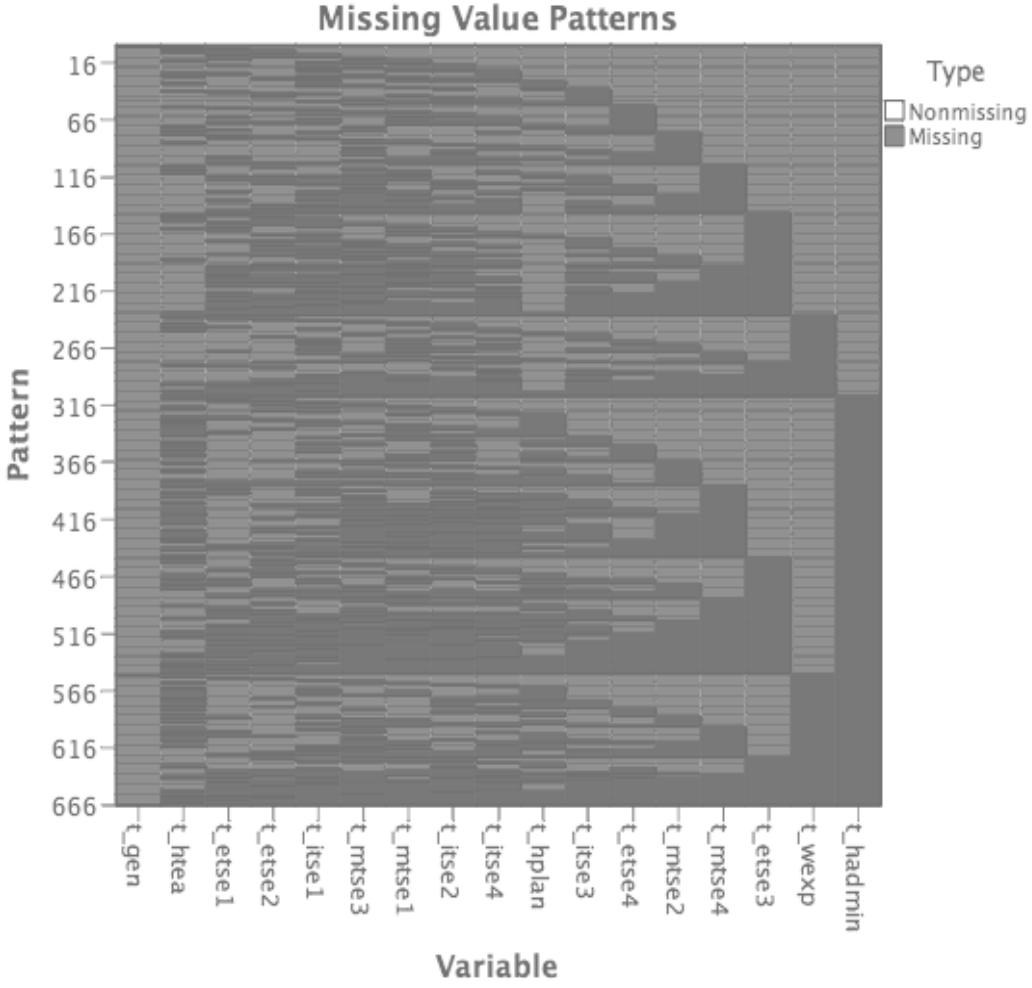
Figure 3-3 Missingness per variables, cases and values



The following table shows that missing values occur systematically. This supports the significant Little's test that revealed that the data are not missing MCAR. The graph below shows which variables are missing together with which other variables and which pattern

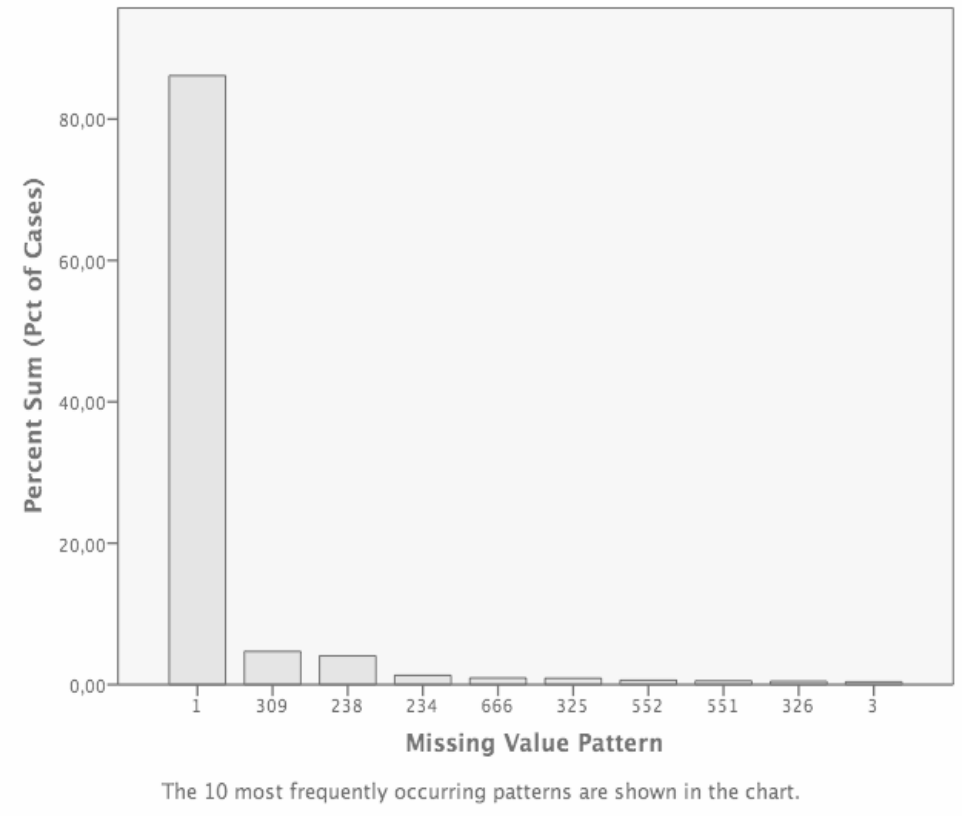
they form. Each pattern belongs to a group of study participants with the same pattern of complete and incomplete cases.

Figure 3-4 Missing value patterns



The bar chart bellows shows how often the ten most frequent missing data patterns occur. The x-axis shows the number of the pattern, the y-axis how often it occurs.

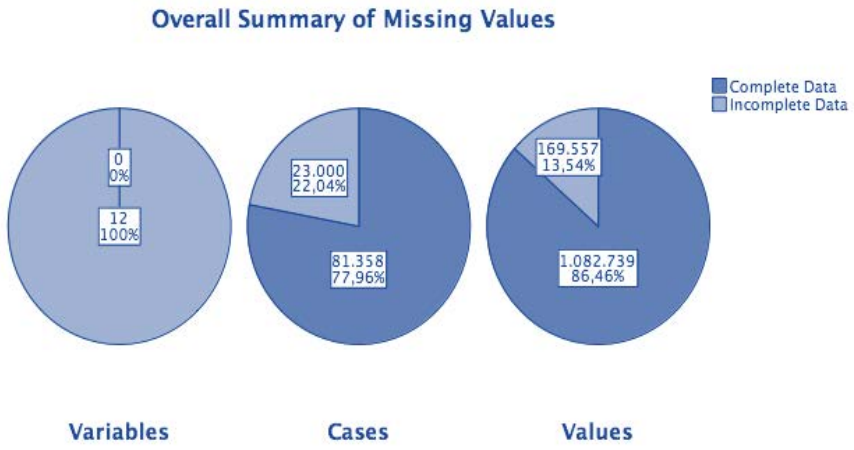
Figure 3-5 10 most frequently occurring patterns



3.10.2 CLASSROOM VARIABLES

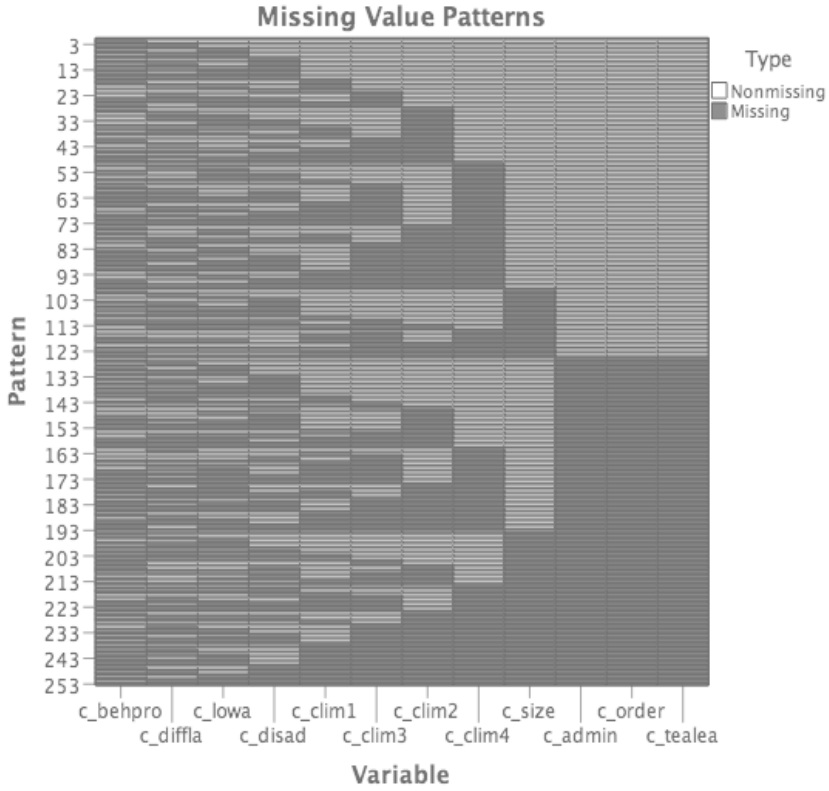
Among the classroom variables, most missing values occurred by cases/respondent and least by variables.

Figure 3-6 Missingness per variables, cases and values



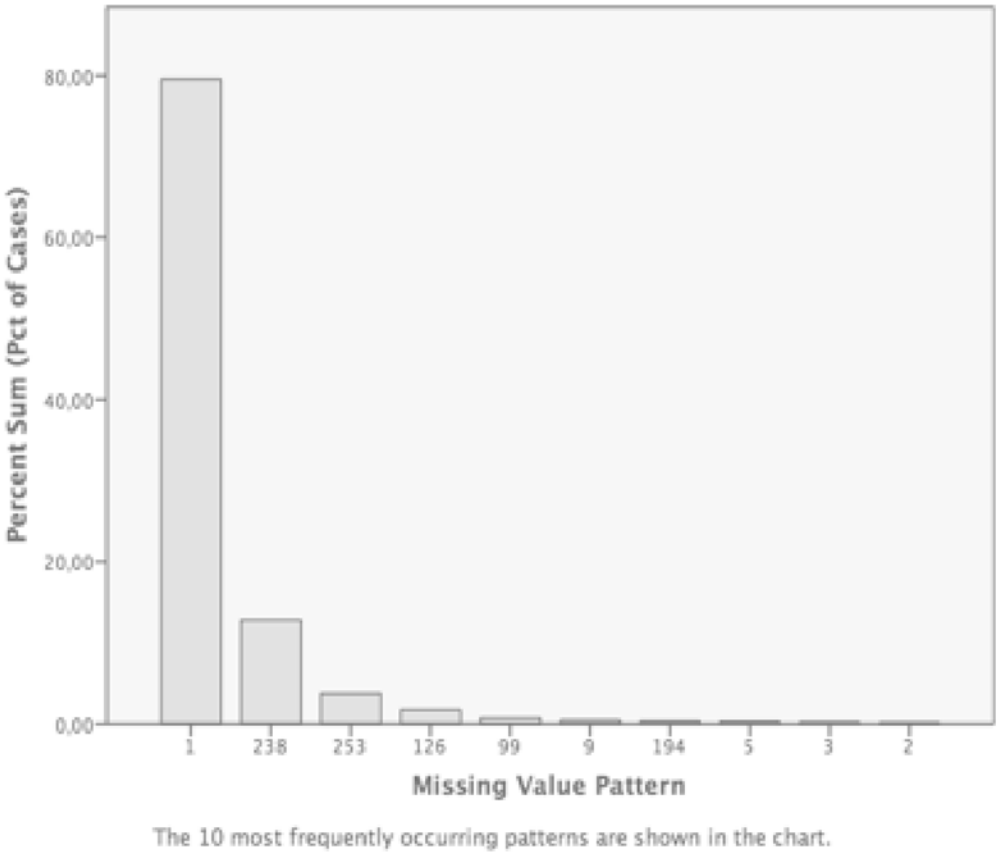
As for the teacher characteristics, the table with missing value patterns shows systematic occurrence of missing data and thus supports the significant Little's test that suggests that the data are not missing MCAR.

Figure 3-7 Missing value patterns



The bar chart shows the ten most frequently occurring missing data patterns. As above, there is one prevailing missing data pattern followed by several less frequent ones.

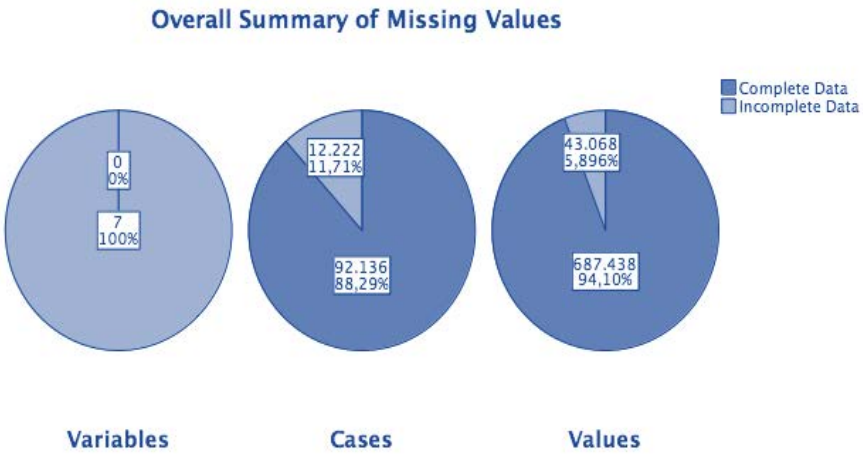
Figure 3-8 10 most frequently occurring patterns



3.10.3 PRINCIPAL VARIABLES

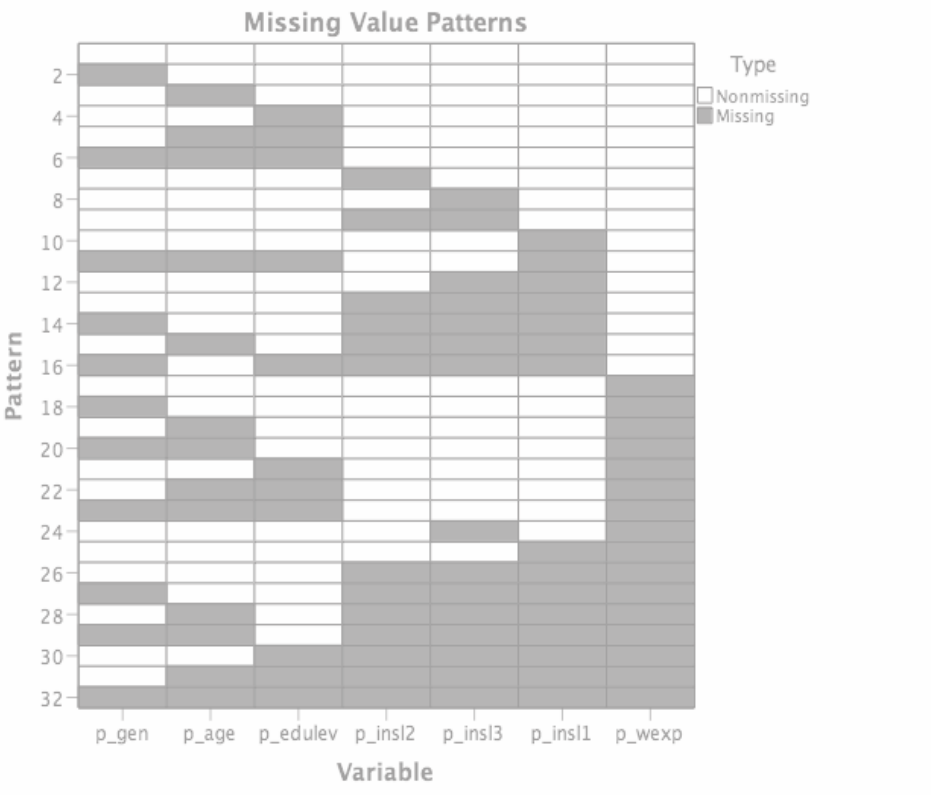
For the principal variables, a similar behaviour is shown as for the classroom variables: Least missing data occurs by variable, most by cases/respondents.

Figure 3-9 Missingness per variables, cases and values



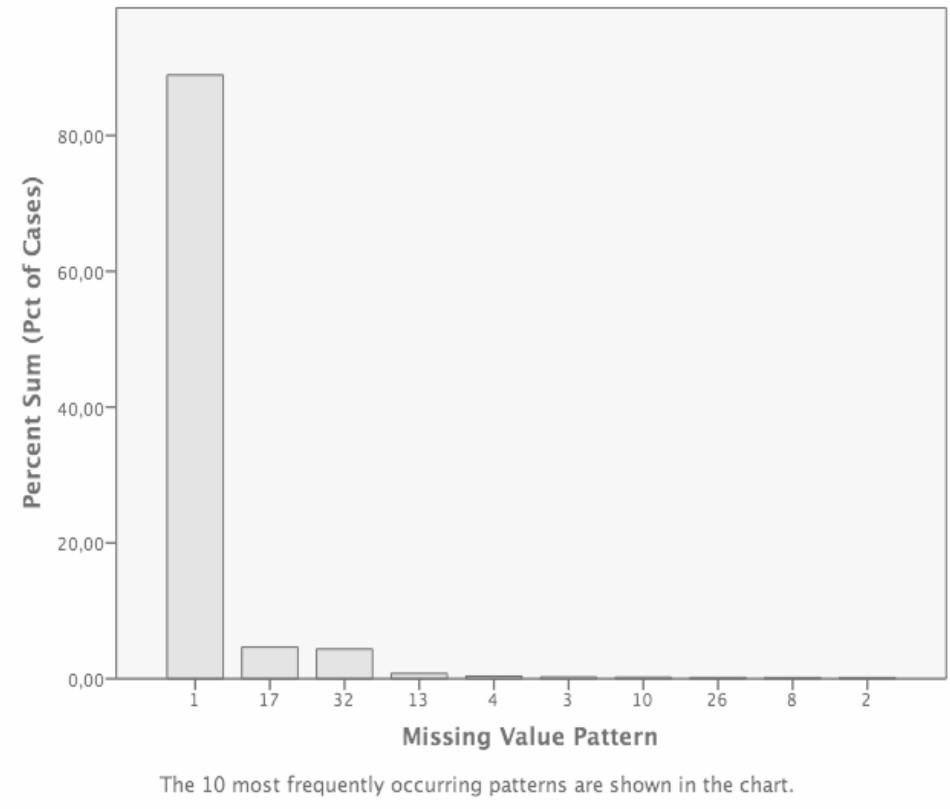
The table below displays the systematic occurrence of missing and existent values for principals. It supports the significant Little's test for missing values that suggests that the data are not missing MCAR.

Figure 3-10 Missing value patterns



The bar charts below show the ten most frequently occurring missing data patterns. There is one prevailing pattern, followed by several severely less frequently occurring ones.

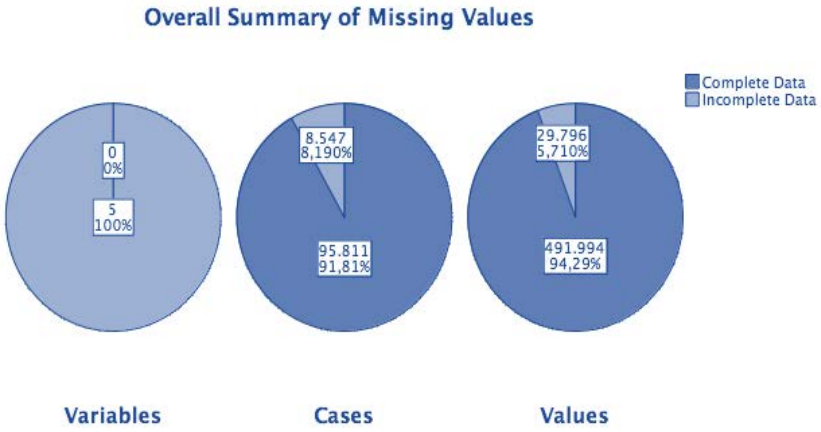
Figure 3-11 10 most frequently occurring patterns



3.10.4 SCHOOL VARIABLES

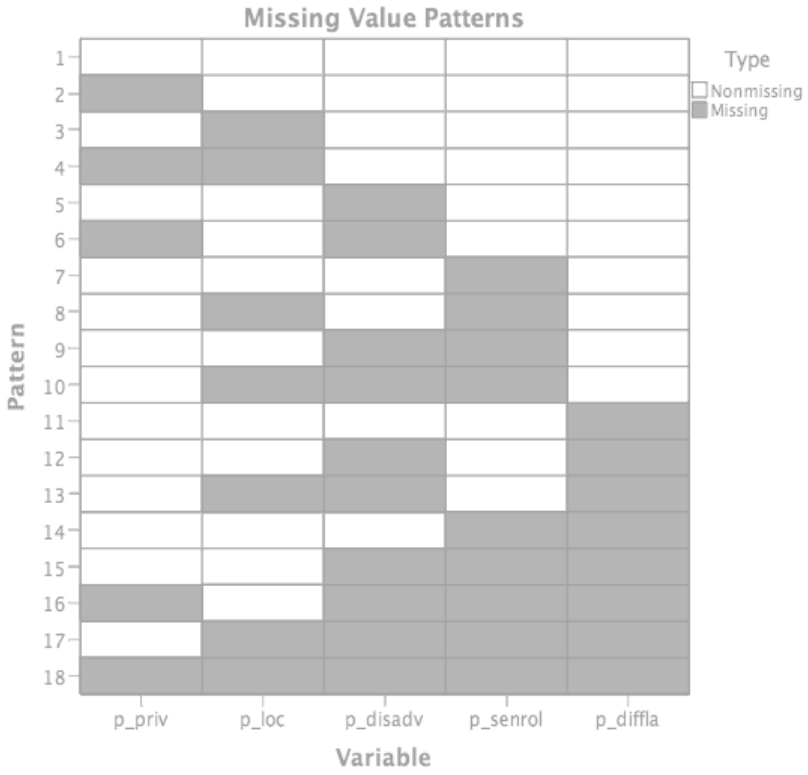
The last set of missing value patterns for school characteristics follows the system of the previous ones: Least missing values occur by variables, most by cases. However, the difference between missingness by cases and missingness by values is smaller than for the other set of teacher, classroom and principal variables.

Figure 3-12 Missingness per variables, cases and values



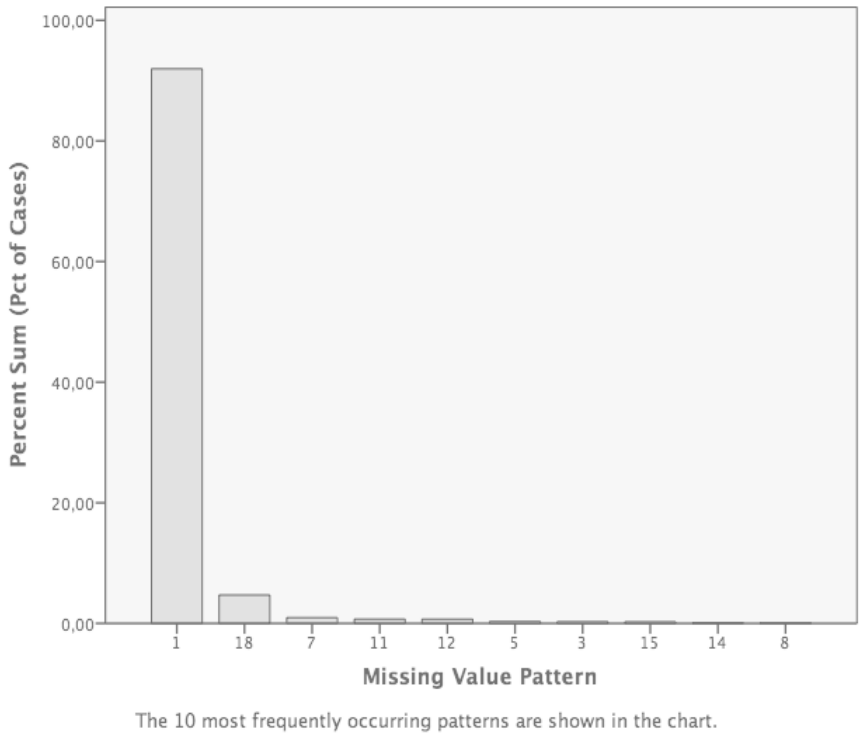
The missing data pattern below, again, supports the significant result of Little's test that suggests that school variables are not missing MCAR (include results from Little's test).

Figure 3-13 Missing value patterns



The bar chart below shows that there is one dominant missing value pattern, and nine less frequently occurring ones.

Figure 3-14 10 most frequently occurring patterns



3.11 SUMMARY

The section above had the following aims: First, it aimed at discussing several ways of treating missing data in the social sciences with respect to underlying reasons for missingness, in order to be able to argue for the appropriate imputation strategy to apply. I chose multiple imputation for the majority of the variables to be included in my thesis, as it yields the most reliable estimates for multilevel data. This imputed data set will be used for both research questions. For the 1st RQ the full data set will used, and for the 2nd research question, the respective countries will be chosen from the main data set and then analysed. Second, the section aimed at describing the imputation strategy used here and thirdly, the missing data patterns were described to provide a data driven rationale for the implementation of the imputation method.

CHAPTER 4 RESULTS: PREDICTING TEACHERS' SELF-EFFICACY BY TEACHER-, CLASSROOM-, PRINCIPAL- AND SCHOOL CHARACTERISTICS

This chapter presents the results for the first set of research questions.

1. To what extent can teacher self-efficacy be explained by teacher characteristics?
2. To what extent can teacher self-efficacy be explained by classroom characteristics?
3. To what extent can teacher self-efficacy be explained by principal characteristics?
4. To what extent can teacher self-efficacy be explained by school characteristics?

Based on previous research (discussed in the literature review), the following hypotheses were also stated and investigated:

1. Male teachers have a higher sense of TSE
2. More experienced teachers have a higher sense of TSE
3. Teachers who teach higher performing students have a higher sense of TSE
4. A warmer and more organised classroom climate is related to a higher sense of TSE
5. Role-model behaviour of the principal (e.g. in instructional leadership) is related to a higher sense of TSE

It is separated into sections according to the three sub-scales of teachers' self-efficacy that are the focus of this research study. The results linked to each of the four research questions noted above are then presented in sub-sections. The three measures of TSE (self-efficacy in instruction, self-efficacy in student engagement and self-efficacy in classroom management) have been analysed in separate multilevel models using a common framework to include the different sets of predictors. The result for each aspect of the three features of TSE are presented and interpreted in each sub-section and then compared in concluding sections.

Tables 18-20 contain the numeric results for the various models that predict the three measures of TSE.

Table 18 Results for TSE in Student Engagement

	Model 0			Model 1				Model 2				Model 3				Model 4			
	β	p	S.E.	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES
Response																			
Fixed Part																			
intercept	3.119	***	.050	3.100	***	.050		3.095	***	.050		3.091	***	.050		3.119	***	.050	
Teacher																			
Gender				.024	***	.003	.10	.026	***	.003	.10	.026	***	.003	.10	.026	***	.003	.10
Educational Level				.003	n.s.	.004	.01	.001	n.s.	.004	.00	.001	n.s.	.004	.00	.000	n.s.	.004	.00
Teaching Exp.				.004	***	.000	.35	.003	***	.000	.26	.003	***	.000	.26	.003	***	.000	.26
Hours Teaching				.002	***	.000	.16	.002	***	.000	.16	.002	***	.000	.16	.002	***	.000	.16
Hours Planning				.002	***	.000	.11	.002	***	.000	.11	.002	***	.000	.11	.002	***	.000	.11
Hours Admin. Work				.003	***	.000	.12	.003	***	.000	.12	.003	***	.000	.12	.003	***	.000	.12
Construct. Bel.				.043	***	.001	.71	.041	***	.001	.68	.041	***	.001	.68	.041	***	.001	.68
Classroom																			
Class size								.001	***	.000	0.08	.001	***	.000	.08	.001	***	.000	.08
Classroom Climate								.033	***	.001	0.75	.033	***	.001	.75	.033	***	.001	.75
Low Student Achievement								-.017	***	.002	-0.14	-.017	***	.002	-.14	-.017	***	.002	-.14
Teaching & Learning								.000	n.s.	.000	0.00	.000	n.s.	.000	.00	.000	n.s.	.000	.00
Keeping Order								-.001	*	.000	-0.10	-.001	*	.000	-.10	-.001	*	.000	-.10
Admin. Tasks								.000	n.s.	.000	0.00	.000	n.s.	.000	.00	.000	n.s.	.000	.00
Different Language								.011	***	.002	0.10	.011	***	.002	.10	.011	***	.002	.10
Disadv. Background								.030	***	.002	0.28	.030	***	.002	.28	.031	***	.002	.29
Principal																			
Gender												.011	*	.004	.05	.010	*	.004	.04
Age												.000	n.s.	.000	.00	.000	n.s.	.000	.00
Educational Level												-.001	n.s.	.000	.00	-.003	n.s.	.006	-.01
Work Experience												.000	n.s.	.006	.00	.000	n.s.	.000	.00
Instructional Leadership												.004	**	.001	.06	.004	**	.001	.06
School																			
Student Enrolment																.000	n.s.	.000	.00
Type of School																-.035	***	.006	-.11
Commuity Size																.006	***	.001	.07
Random Part																			
Country	.070	***	.018	.067	***	.017		.068	***	.018		.067	***	.017		.067	***	.017	
School	.006	***	.000	.005	***	.000		.004	***	.000		.004	***	.000		.004	***	.000	
Teacher	.159	***	.002	.150	***	.002		.142	***	.001		.142	***	.001		.142	***	.001	
-2*loglikelihood:	742,793			738,583				734,261				734,237				734,178			
Δ -2*loglikelihood				-4,210				-4,321				-24				-60			
% variance			Explained variance			Explained variance				Explained variance				Explained variance					
Country	29.8%		Country	4.3%		Country	2.9%			Country	4.3%			Country	4.3%				
School	2.6%		School	16.7%		School	33.3%			School	33.3%			School	33.3%				
Teacher	67.7%		Teacher	5.7%		Teacher	10.7%			Teacher	10.7%			Teacher	10.7%				
ICC																			
Country	29.7%	***		30.2%	***			31.9%	***			31.6%	***			31.6%	***		
School	2.5%	***		2.4%	***			1.9%	***			1.9%	***			1.9%	***		
Teacher	67.8%	***		67.4%	***			66.2%	***			66.5%	***			66.6%	***		

Table 19 Results for TSE in Instruction

	Model 0			Model 1				Model 2				Model 3				Model 4			
	β	p	S.E.	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES
Response																			
Fixed Part																			
intercept	3.189	***	.047	3.147	***	.047		3.142	***	.046		3.160	***	.046		12.267	***	1.380	
Teacher																			
Gender				.043	***	.002	.32	.045	***	.002	.34	.045	***	.002	.34	.045	***	.002	.34
Educational Level				.008	*	.003	.05	.007	*	.003	.04	.007	*	.003	.04	.006	*	.003	.04
Teaching Exp.				.002	***	.000	.33	.001	***	.000	.17	.001	***	.000	.17	.001	***	.000	.17
Hours Teaching				.002	***	.000	.29	.001	***	.000	.15	.001	***	.000	.15	.001	***	.000	.15
Hours Planning				.002	***	.000	.21	.002	***	.000	.21	.002	***	.000	.21	.002	***	.000	.21
Hours Admin. Work				.003	***	.000	.22	.003	***	.000	.22	.003	***	.000	.22	.003	***	.000	.22
Construct. Bel.				.032	***	.001	.99	.031	***	.001	.96	.031	***	.001	.96	.031	***	.001	.96
Classroom																			
Class size								.001	***	.000	0.16	.001	***	.000	.16	.001	***	.000	.16
Classroom Climate								.022	***	.001	0.95	.022	***	.001	.95	.022	***	.001	.95
Low Student Achievement								-.003	*	.001	-0.05	-.003	*	.001	-.05	-.003	*	.001	-.05
Teaching & Learning								.000	n.s.	.000	0.00	.000	n.s.	.000	.00	.000	n.s.	.000	.00
Keeping Order								.000	n.s.	.000	0.00	.000	n.s.	.000	.00	.000	n.s.	.000	.00
Admin. Tasks								.000	n.s.	.000	0.00	.000	n.s.	.000	.00	.000	n.s.	.000	.00
Different Language								.009	***	.001	0.16	.009	***	.001	.16	.008	***	.001	.14
Disadv. Background								.023	***	.001	0.41	.023	***	.001	.41	.023	***	.001	.41
Principal																			
Gender												.007	*	.003	.06	.007	*	.003	.06
Age												.000	n.s.	.000	.00	.000	n.s.	.000	.00
Educational Level												.001	n.s.	.005	.00	.000	n.s.	.005	.00
Work Experience												.000	n.s.	.000	.00	.000	n.s.	.000	.00
Instructional Leadership												.004	***	.001	.12	.004	***	.001	.12
School																			
Student Enrolment																.000	n.s.	.000	.00
Type of School																-.004	n.s.	.004	-.02
Commuity Size																.004	***	.001	.09
Random Part																			
Country	.027	***	.007	.026	***	.007		.026	***	.007		.026	***	.007		.026	***	.007	
School	.003	***	.000	.003	***	.000		.003	***	.000		.003	***	.000		.003	***	.000	
Teacher	.095	***	.001	.091	***	.001		.089	***	.001		.089	***	.001		.089	***	.001	
-2*loglikelihood:	728,517			724,327				721,229				721,190				721,174			
Δ -2*loglikelihood				-4,190				-3,099				-38				-16			
% variance				Explained variance				Explained variance				Explained variance				Explained variance			
Country	21.6%			Country	3.7%			Country	3.7%			Country	3.7%			Country	3.7%		
School	2.4%			School	n.s.			School	n.s.			School	n.s.			School	n.s.		
Teacher	76.0%			Teacher	4.2%			Teacher	6.3%			Teacher	6.3%			Teacher	6.3%		
ICC																			
Country	21.6%	***		21.8%	***			22.2%	***			21.8%	***			22.0%	***		
School	2.4%	***		2.3%	***			2.2%	***			2.2%	***			2.1%	***		
Teacher	76.0%	***		75.9%	***			75.6%	***			76.0%	***			75.9%	***		

Table 20 Results for TSE in Classroom Management

	Model 0			Model 1				Model 2				Model 3				Model 4			
	β	p	S.E.	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES	β	p	S.E.	ES
Response																			
Fixed Part																			
intercept	3.305	***	.037	3.289	***	.038		3.283	***	.038		3.284	***	.037		3.297	***	.038	
Teacher																			
Gender				.023	***	.004	.07	.025	***	.004	.08	.025	***	.004	.08	.026	***	.004	.08
Educational Level				-.011	*	.005	-.03	-.016	**	.005	-.04	-.016	**	.005	-.04	-.016	**	.005	-.04
Teaching Exp.				.005	***	.000	.35	.004	***	.000	.28	.004	***	.000	.28	.004	***	.000	.28
Hours Teaching				.003	***	.000	.19	.002	***	.000	.13	.002	***	.000	.13	.002	***	.000	.13
Hours Planning				-.001	***	.000	-.04	-.001	***	.000	-.04	-.001	***	.000	-.04	-.001	***	.000	-.04
Hours Admin. Work				.005	***	.000	.16	.005	***	.000	.16	.005	***	.000	.16	.005	***	.000	.16
Construct. Bel.				.029	***	.001	.38	.026	***	.001	.34	.026	***	.001	.34	.026	***	.001	.34
Classroom																			
Class size								.002	***	.000	0.13	.002	***	.000	.13	.002	***	.000	.13
Classroom Climate								.054	***	.001	0.99	.054	***	.001	.99	.054	***	.001	.99
Low Student Achievement								.022	***	.002	0.15	.022	***	.002	.15	.022	***	.002	.15
Teaching & Learning								.001	n.s.	.000	0.11	.001	n.s.	.000	.11	.001	n.s.	.000	.11
Keeping Order								-.001	***	.000	-0.08	-.001	***	.000	-.08	-.001	***	.000	-.08
Admin. Tasks								.001	n.s.	.000	0.05	.001	n.s.	.000	.05	.001	n.s.	.000	.05
Different Language								.001	n.s.	.002	0.01	.001	n.s.	.002	.01	.001	n.s.	.002	.01
Disadv. Background								.025	***	.002	0.19	.025	***	.002	.19	.025	***	.002	.19
Principal																			
Gender												-.003	n.s.	.004	-.01	-.002	n.s.	.004	-.01
Age												.000	n.s.	.000	.00	.000	n.s.	.000	.00
Educational Level												-.008	n.s.	.007	-.02	-.009	n.s.	.007	-.02
Work Experience												.000	n.s.	.000	.00	.000	n.s.	.000	.00
Instructional Leadership												.003	*	.001	.04	.003	*	.001	.04
School																			
Student Enrolment																.000	n.s.	.000	.00
Type of School																-.016	*	.007	-.04
Commuity Size																-.001	n.s.	.002	-.01
Random Part																			
Country	.040	***	.010	.041	***	.010		.042	***	.011		.041	***	.010		.040	***	.010	
School	.007	***	.000	.007	***	.000		.005	***	.000		.005	***	.000		.005	***	.000	
Teacher	.247	***	.002	.241	***	.002		.224	***	.002		.224	***	.002		.224	***	.002	
-2*loglikelihood:	672,867			670,525				663,380				663,369				663,360			
Δ -2*loglikelihood				-2,342				-7,145				-11				-8			
% variance				Explained variance				Explained variance				Explained variance				Explained variance			
Country	13.6%			Country	n.s.			Country	n.s.			Country	n.s.			Country	n.s.		
School	2.4%			School	n.s.			School	28.6%			School	28.6%			School	28.6%		
Teacher	84.0%			Teacher	2.4%			Teacher	9.3%			Teacher	9.3%			Teacher	9.3%		
ICC																			
Country	13.7%	***		14.2%	***			15.4%	***			15.2%	***			14.9%	***		
School	2.2%	***		2.3%	***			1.8%	***			1.8%	***			1.8%	***		
Teacher	84.0%	***		83.5%	***			82.8%	***			8.3%	***			83.2%	***		

4.1 TEACHER'S SELF-EFFICACY IN STUDENT ENGAGEMENT (ETSE)

This section contains the results and interpretations of these results of the research question, how teacher-, classroom-, principal- and school characteristics are related to teacher's self-efficacy in student engagement. It is structured according to the four sub-questions investigating predictors on each level and contains a summary at the end.

4.1.1 BASELINE MODEL

The baseline (null) model contains no predictors and serves to partition the variance in ETSE between different levels (country, school, teacher). It is mainly used as a means to compare the way the inclusion of different sets of predictors affect variance in subsequent models. These will contain predictors concerning the teachers themselves, the classroom in which teaching takes place, the principal responsible for the school the teachers work in and finally predictors that characterise the school in which the teachers work. The reduction in unexplained variance will be identified and model fit statistics compared.

For self-efficacy in student engagement (ETSE), the initial (null) model (0) -2log likelihood was 742,793 and it revealed that 29.8% of the variance occurred among countries, 2.6% among schools and 67.7% among teachers. This reveals the relative importance of different sources of variation in the nested data prior to adding any potential predictors.

4.1.2 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY TEACHER CHARACTERISTICS?

In the first model (1), the variables teacher gender (female), educational level, teaching experience, weekly hours spent on teaching, planning lessons and administrative work as well as to what extent the teacher holds constructivist beliefs about teaching are included as potential predictors of self-efficacy in student engagement.

The model (1) fit improved compared to the baseline model (0), as the -2log likelihood decreased to 738.583 (**Δ** = -4.210). By adding explanatory variables on the teacher level, 5.7% of the variance could be explained, 16.7% among schools and 4.3% among countries.

All factors except the teacher's educational level show a significant positive association with ETSE: Female gender (**β** = .024, p < .001, ES = .10), teaching experience (**β** = .004, p < .001, ES = .35), hours per week spent on teaching (**β** = .002, p < .001, ES = .16), hours per week spent on planning lessons (**β** = .002, p < .001, ES = .11), hours spent on administrative work (**β** = .003, p < .001, ES = .12) and constructivist beliefs about teaching (**β** = .043, p < .001, ES = .71). With the exception of the measure of constructivist beliefs about teaching, which had a medium effect size[8], all other predictors had only a very small to small effect on ETSE.

The result that female teachers feel more self-efficacious than male teachers contradicts the hypothesis derived from previous research, where male teachers tended to have a higher sense of SE (i.e. Cousins et al. 1996); although the effect size is only small (ES = .10). However, previous results are mixed, as there is also support for a higher level of TSE for female teachers reported in the literature, as, for instance, by Coladarci (1992) or Fackler & Malmberg (2016) or no gender effect (Malmberg et al., 2014). In the same way, the result that more experienced teachers feel more self-efficacious is partly in line with previous research, which has shown mixed results: a non-linear relationship where TSE first rose and then dropped (Burke & Spero, 2005; Klassen & Chiu, 2010; Pas et al. 2012), but also a positive association for work experience and ETSE (Giallo & Little, 2003; Wolters & Daugherty, 2007; Cousins et al., 1996; Fackler & Malmberg, 2016). Gkolia (2014) found that more experienced teachers do not feel more self-efficacious in student engagement than less experienced teachers. However, based on these mixed results, (Zee & Koomen, 2016) synthesised that apparently the positive effects of accumulated experience do not seem to last over the course of the career. It seems to be noteworthy that despite the mixed results of previous research, the effect size with ES = .35 in this study is of medium strength and the predictor of rather high importance.

[8] Strength of effect sizes according to Tymms, 2004 & Marsh et al., 2009, for multilevel models (comparable with Cohen's *d*):

<.3 small effect size

<.5 medium effect size

>.8 big effect size

Moreover, it is possible that teachers, who are expecting to succeed in engaging their students, not only spend more time per week on planning their lessons, but also on teaching as they might expect their efforts to be worthwhile. Alternatively spending more time on planning and teaching may help to boost teachers' self-efficacy in terms of engaging their students, or the associations may be reciprocal (but no data exist in the cross-sectional survey to test this possibility). The effect sizes, however, were both only rather small with ES = .16 and ES = .11. Mastery experience (Bandura, 1977) of prior teaching of that group could serve as a potential source for self-efficacy here and is in line with previous research as hypothesised. Constructivist beliefs about teaching, then, comprise the attitude towards the students as active learners. Holzberger et al. (2014) reported that teachers with a high level of TSE better engage students in the lessons, which seems to be in accordance with this finding. Teachers, who in turn act according to the conviction that students need to be involved in the lesson in order to successfully learn, might well feel able in doing so. This might also reflect the teachers' aim of enabling the student and pushing their boundaries regardless of their upbringing or social background. Constructivist beliefs about teaching revealed the comparatively largest effect size with ES = .71, of all teacher characteristics. Hence, the predictor shows high importance.

4.1.3 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY CLASSROOM CHARACTERISTICS?

In the second part of this research question, classroom-related variables were added to the model (2). These are class size, classroom climate, low student achievement, percentage of time in the classroom spent on teaching and learning, keeping order and doing administrative tasks, the proportion of students who speak a different mother tongue from the language of instruction and the proportion of students who come from a disadvantaged background. The last two of these measures can be viewed as contextual indicators that might shape the classroom environment with respect to inter-individual socio-economic differences of the students.

The model (2) fit improved again compared to the first model (1), as the 2log likelihood now decreased 734,261 (**Δ** = -4,321). Among countries, 2.9% of the variance could now be explained, 33.3% among schools and 10.7% among teachers.

Adding these classroom characteristics did not significantly change the associations of teacher characteristics and ETSE found in model 1. From the newly added classroom predictors, class size (**β** = .001, p < .001, ES= .08) is positively related to ETSE, as well as a good classroom climate (**β** = .033, p < .001, ES= .75), which is in line with previous research (Ashton et al., 1983; Künsting et al., 2016). This classroom climate measure is based on items indicating that the students care about a good classroom environment or make an effort to follow the teacher's rules and revealed the largest effect size of all classroom predictors, whereas the class size only had a rather small effect size. In contrast, low average student achievement (**β** = -.017, p < .001, ES = -.14), shows a negative relationship with ETSE and has a small effect size, indicating that this contextual feature predicts lower ETSE. This is in line with previous research (Ashton et al., 1983; Malmberg et a., 2014; Fackler & Malmberg, 2016). Spending time on keeping order in the classroom also shows a negative association with ETSE (**β** = -.001, p < .01, ES = -.10), whereas the time spent on administrative tasks and teaching and learning were not significantly associated with a high sense of TSE in student engagement. But, the percentage of students with a different mother tongue from the language of instruction (**β** = .011, p < .001, ES = .10) and the percentage of students from a disadvantaged background (**β** = .030, p < .001, ES = .28), in contrast, are positively related to a teacher's self-efficacy in student engagement. In all three cases, the effect sizes were only small, indicating that they do not make a large difference in ETSE.

Apart from classroom climate, which has a medium effect size (.75), all other effect sizes were small (< .5). It should be noted that such small effects sizes are common in school and teacher effectiveness studies (e.g. Scheerens et al., 2013).

As major source for many of the relationships of ETSE with classroom predictors, mastery experienced, or put more simply, having success, seems to be worthwhile exploring. Low stu-

dent achievement, good classroom climate and little time spent on keeping order in the classroom may all provide sources of feedback for the teacher on how well the students are following the lesson, in terms of how deeply they engage in their classroom activities that should help support their learning. Attentive students should be less likely to interrupt lessons through misbehaviour and thus the classroom climate would tend to be better and the teacher needs to spend less time on keeping order. In contrast, badly performing students might make the teachers question their efficacy in terms of creating lessons the students are engaged in.

Furthermore, the positive association of ETSE with the percentage of students from disadvantaged backgrounds and a different mother tongue than the language of instruction could be viewed as surprising results, as this has not yet been shown in previous research on TSE (i.e. Geerlings et al., 2018), but nonetheless are features that may well shape the immediate work environment of the teacher. Both effect sizes, however were rather small with ES = .10 for students with a different language and ES = .28 for students from a disadvantaged background; although the latter effect size almost reaches medium strengths. For example, Strand (2012) found that contextual measures of the environment (in terms of composition of the school's student intake) reveal an association between student ethnicity and education outcomes, where low SES students with migrant background outperform local (defined as 'white British') students. In this study, parental academic aspirations served as a possible explanation for this result, which were higher for parents with migrant backgrounds than those of local low SES parents. However, rather the opposite could have been expected here, as language is the main means through which teaching takes place and teachers who may find communication with their students more challenging in high non-mother tongue contexts, might not be expected to perceive that they can have the same success with students for whom there may be language barriers to learning. But it is possible that teachers in high disadvantage contexts with more minority students may feel that their contribution has greater potential for positive effects because such bilingual/multilingual students may have much improved outcomes if they develop fluency in the language of instruction.

4.1.4 To what extent can variation in ETSE be explained by principal characteristics?

In the third model, also, variables related to the school principal were added to investigate the possibility of associations with principal characteristics and ETSE. These are namely female gender of the principal, age, educational level, work experience and an instructional style of leadership.

Compared to the previous model (2), the -2log likelihood again decreased to 734,237 (**Δ** = -24), and among countries (4.3%), schools (33.3%) and teachers (10.7%) of the variance could be explained.

When adding principal characteristics, none of the previously added teacher or classroom predictors changed their relationship with ETSE significantly.

From the principal characteristics, only two showed positive associations related to ETSE: Female gender (**β** = .011, p < .01, ES = .05) and an instructional style of leadership (**β** = .004, p < .05, ES = .06). The measure of instructional leadership is based on items linked to the principal's role model behaviour and strong ideas about the school's academic focus. Thus, it can be seen as a source of additional influence on school culture and organisational aims from which the teachers might be able to benefit. The effect sizes for these two predictors are only very small, however the instructional leadership measure is in the direction that would be hypothesised from research on school leadership (i.e. Coladarci et al 1992; Raudenbush et al 1992; Skaalvik & Skaalvik, 2007; Guo et al, 2011).

The positive association of female gender of the principal, however, is a new result, which has not been shown so clearly in association with high teacher self-efficacy so far. The only study finding a higher sense of SE for female teachers is by Colardarci (1992) in relation to teaching commitment. However, this study does not investigate indirect effects on TSE as previous studies finding this result did. Thus, reasoning about this does not seem appropriate. To explore these relations, more studies into the gender effect and TSE would be needed. Instruc-

tional style of leadership, on the contrary, and TSE have been explored beforehand (i.e. Coladarci et al. 1992, Raudenbush et al. 1992, Guo et al. 2012; Fackler & Malmberg, 2016) and the hypothesis that TSE and instructional leadership style have a positive association could be confirmed. Given the features of the instructional leadership style like goal-setting, academic focus of the school and role model behaviour of the principal, the teachers might well find themselves in a supportive environment, where the behaviour of the principal supports teachers' own mastery experience through alignment with the teachers' own attempts to engaging students in the lessons to support learning and achievement.

4.1.5 TO WHAT EXTENT CAN VARIATION IN ETSE BE EXPLAINED BY SCHOOL CHARACTERISTICS?

In the final model (4), three school characteristics were added to the analysis: number of students enrolled at school, school type (private versus public) and community size. Among the teacher and classroom related characteristics, none significantly changed in their effects after adding school characteristics to the analysis.

The -2log likelihood again improved as it became 734,178 (**Δ** = -60) and the amount of variance explained among countries was 4.3%, among schools 33.3% and among teachers 10.7%, the latter was unchanged from the previous model (3).

Private schools were negatively related to ETSE (**β** = -.035, $p < .001$, ES = -.11) and community size, positively (**β** = .006, $p < .001$, ES = .07). Both results are contrary to previous research (Knoblauch & Wolfoolk-Hoy, 2008; Fackler & Malmberg, 2016). These are both contextual factors that may help shape the school environment in which the teachers find themselves. Private schools are often chosen very carefully by students' parents and thus the composition of students is likely to be different from public schools, which tend to have a more diverse and less selected student group (i.e. Yun & Reardon, 2005). The size of the community provides an indicator for bigger cities. In most countries, there is an urban-rural divide in educational achievement and typically urban areas have more educated populations. This contextual difference could explain the positive association with ETSE because there may be a more positive attitude to education among the parent and student population linked to better employment opportunities in urban areas for more qualified young people (Copus et al., 2006).

As in the previous model, the effect sizes for the school predictors are only small and thus overall hint at those school characteristics not contributing to a large difference in ETSE in the 32 OECD countries.

4.2 TEACHERS' SELF-EFFICACY IN INSTRUCTION (ITSE)

This section contains the results and interpretations of these results of the research question, how teacher-, classroom-, principal- and school characteristics are related to teacher's self-efficacy in instruction, which will draw on the effect sizes of the predictors. It is structured according to the four sub-questions investigating predictors on each level and contains a summary at the end.

4.2.1 BASELINE MODEL

The baseline model (0) for self- efficacy in instruction (ITSE) again serves as the initial comparison for the following models, where explanatory variables will be added, firstly related to the teacher, secondly the classroom environment, thirdly the principal and fourthly the school in which the teachers work.

For the baseline model for ITSE, the initial -2log likelihood was 728,517 and 21.6% variance was found among countries, 2.4 % among schools and 76.0% among teachers.

4.2.2 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY TEACHER CHARACTERISTICS?

In the first model (1), teacher gender (female), educational level, teaching experience, weekly hours spent on teaching, planning lessons and administrative work as well as to which extent the teacher holds constructivist beliefs about teaching are analysed as potential predictors of ITSE.

After adding teacher related variables, the model (1) fit improved with a drop in the -2loglikelihood to 724,327 (**Δ** = -4,190). Compared to the baseline model (0), among countries 3.7% of the variance could be explained, 4.2% among teachers but none of the variance among schools.

All predictors added in model (1) show a significant relationship with self-efficacy in instruction. Female gender (**β** = .043, p < .001, ES = .32), a higher educational level (**β** = .008, p < .01, ES = .05) and years of teaching experience (**β** = .002, p < .001, ES = .33). Teachers' female gender is again contrary to the hypothesis, especially with a medium-size effect size, but the positive association with work experience supports the hypothesis that more experienced teachers feel more self-efficacious and the effect size supports this (i.e. Coladarci, 1992; Chester & Beaudin, 1996; Fackler & Malmberg, 2016; Cousins et al., 1996). Both a higher educational level and the years of experience in teaching, could, again, hint at a source of mastery experience that the teacher gains with time studying and spending in the classroom; although the effect size of the teachers' educational level is only very small and thus lends less support to the hypothesis than the teaching experience gained over the years, with respect to teacher gender and age, results found in previous research are mixed.

In the same way as identified in the previous analyses of ETSE, spending more hours on teaching (**β** = .002, p < 001, ES = .29), hours planning (**β** = .002, p < .001, ES = .21) and hours doing administrative work (**β** = .003, p < .001, ES = .22) show a positive association with ITSE and the effect sizes are small, but the ES for hours spent on teaching is considerably close to the .3-threshold. These three types of time allocation of the teacher might hint that those teachers who feel they are capable of teaching their students are not only spending more time on preparing lessons and focus on teaching as such, but also feel more capable of doing general work that is associated with the teaching profession, like administrative work. This, in turn, may mean teachers are more likely to believe that they are able to influence how well students learn. Investigating the teacher's time allocation, however, is novel, but could be interpreted as a positive attitude towards the teacher's workload: high TSE teachers are putting more effort into their job, as they might expect to succeed. It is most notable that the effect size for the hours spent on teaching, which is conceptually most closely related to ITSE, is the comparatively highest one.

A strong relationship and big effect size, which is the largest among all teacher predictors, was found with constructivist beliefs about teaching (**β** = .032, p < .001, ES = .99). The measure of constructivist beliefs about teaching concerns the belief that students can be taught to

master and direct their own learning process. Thus, teachers who hold that belief and structure their lessons accordingly, display in their teaching that they believe they are able to overcome external factors such as a student's lack of interest in the subject and can make a difference in their students' learning. This finding is in line with that found for ETSE but the association is stronger in terms of effect size. This is in line with previous research, which has found that high TSE teachers tend to actively support student learning (Emmer & Hickman, 1991), but also use more challenging cognitive tasks (Li & Liu, 2000).

In the first model, all effect sizes are small according to its classification, except for the predictor constructivist beliefs about teaching (.99). But is has to be mentioned that the range that covers small effect sizes with .05-.33 is rather wide.

4.2.3 To what extent can variation in ITSE be explained by classroom characteristics?

In the second part of this research question investigating self-efficacy in instruction, classroom-related variables were added to the model. These are class size, classroom climate, low student achievement, percentage of time in the classroom spent on teaching and learning, keeping order and administrative tasks, the proportion of students who speak a mother tongue which is different from the language of instruction, and the proportion of students who come from a disadvantaged background. As noted previously, most of these measures (i.e. low student achievement, different mother tongue or students from a disadvantaged background) can be viewed as contextual indicators.

Compared to the first model, the model fit after adding classroom predictors (model 2) improved to 721,229 (**Δ** = -3,099). Among countries 3,7% and teachers 6,3%, of the variance could be explained significantly, but none among schools.

Five of the classroom characteristics that were found to be significantly associated with ITSE: Class size has a low positive relationship and the effect size is also rather small (**β** = .001, $p < .001$, ES = .16), which is in line with previous research (Coladarci, 1992; Raudenbush et al., 1992). A good classroom climate shows a strong positive association and has a big effect size

(**β** = .022, p < .001, ES = .95), as has been hypothesised (i.e. Aslan, 2015). This suggests that in classrooms where there is a positive atmosphere, teachers can experience both mastery in their efforts to teach students and verbal persuasion, as the students may be seen to support their teachers' efforts by taking care of a good classroom environment themselves. This, in turn, could elicit an environment which is supportive of teaching. A classroom context of low student achievement, in contrast, shows a low negative association (**β** = -.003, p < .01, ES = -.05) with ITSE, but the effect size is rather tiny, and is in line with the hypothesis that high TSE teachers have high performing students, whereas the contextual measure of a high proportion of students with a different mother tongue than the language of instruction shows a low positive association and a small effect size (**β** = .009, p < .005, ES = .16). Thus, out of these two contextual classroom measures, the mother tongue seems to be a better predictor. In addition, the proportion of students from a disadvantaged socio-economic background also showed a positive effect (**β** = .023, p < .001, ES = .41), which is of medium size and thus shows the strongest association with ITSE of all three contextual measures. The association between low student achievement and low ITSE could suggest that when students perform badly, teachers can experience a setback in their mastery experience. This then may be associated with a lower sense of their abilities to instruct students well, but due to the very small effect size, this conclusion is only drawn cautiously. This is supported by Fackler & Malmberg, 2016; Malmberg et al. (2014) and Ashton et a. (1983), who found positive associations between TSE and student achievement. Somewhat surprising, however, is the positive association with ITSE and the students' different mother tongue as it could be expected that teachers face greater challenges in supporting student learning in these situations (i.e. Cajkler & Hall, 2009), but again, the effect size is rather small. Language can be seen as the main means of communication and thus instructing students. This is why a high number of students with a different mother tongue could rather be expected to occur with a low perception of ITSE. The opposite was found by Geerlings et al. (2018), who reported a lower level of ITSE in cases of a larger number of minority students in the classroom. Nonetheless, Cajkler & Hall found in their 2009 study that pre-service teacher training for teaching students with different mother tongues, as well as placements and supervisions, contribute to teachers feeling prepared for these situations. It may also be that teachers find that bilingual students can do well and that this

supports positive feelings of ITSE because teachers see language skills as learnable, but this needs to be investigated in future research.

Finally, none of the measures of time allocations (teaching and learning, keeping order and spending time on administrative tasks) show a statistically significant relationship with ITSE in contrast to findings for ETSE. It was hypothesised that the time spent on teaching and learning might have shown a positive effect as it might provide a source of mastery experience for the teachers, when they are able to predominantly perform teaching in the classroom. Teaching is the predominant action to be performed in the classroom and both successfully teaching with the belief to be able to engage students in the lessons and to be able to deliver successful instruction seem to be reasonably reflected by the time teachers actually spend on doing it compared to other duties in the classroom.

The relationship between ITSE and the teacher predictors changed somewhat after including classroom measures, but they remained highly significant. It is worth pointing out that all **β** decreased as did effect sizes (i.e. for constructivist beliefs about teaching .032***->.031***; for teaching experience .002***->.001***) but increased for female gender (.043***->.045***). It appears that there may be some shared variance, classroom factors may be shaping the role of teacher characteristics in predicting ITSE, perhaps via mediating effects. This was not evident in the earlier results for ETSE.

4.2.4 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY PRINCIPAL CHARACTERISTICS?

In the third model, variables describing the principal were also added to investigate associations with principal characteristics and ITSE. These are: female gender of the principal, age, educational level, work experience and an instructional style of leadership.

After adding principal related variables to the analysis, none of the teacher- and classroom predictors changed their association with ITSE. Only two principal characteristics were significantly positively related to higher level of teacher self-efficacy in instruction, but their effect sizes were tiny, respectively small only: female gender (**β** = .007, $p < .01$, ES = .06) and an instructional style of leadership (**β** = .004, $p < .001$, ES = .12). The leadership association is as

has been hypothesised (Fackler & Malmberg, 2016; Guo et al., 2012) but the effect size of female gender is too small to draw conclusions, especially considering the lack of previous research on this topic. As one feature that the instructional leadership style comprises is a focus on academic achievement, teachers working with a principal focusing on that, could experience a more supportive environment for achieving good academic results and thus feel more self-efficacious. Again, the results accord with those for ETSE.

By adding principal characteristics, the model fit was again improved, now showing a -2log likelihood = 721,190 (**Δ** = -38). As for the variance among schools, no more could be explained significantly, but there was more explained among countries 3.7% and among teachers 6.3%. This suggest principal effects can operate at these levels (not only, as might be anticipated, at the school level).

4.2.5 TO WHAT EXTENT CAN VARIATION IN ITSE BE EXPLAINED BY SCHOOL CHARACTERISTICS?

In the fourth model, also school characteristics were added to the analysis: The number of students enrolled at school, the school type (private versus public) and the community size.

The model fit was again slightly improved (-2log likelihood = 721,174; **Δ** = -16). Compared to the previous model, the same amount of variance could be explained among countries (3.7%), and teachers (6.3%), but none among schools.

Adding school characteristics to the analyses led to no significant change in the previous models. However, community size shows a small significant, positive relationship with ITSE and has a very small effect size (**β** = .004, $p < .001$, ES = .09). As already assumed for self-efficacy in student engagement, parents with a high educational level tend to work in bigger cities due to the job opportunities there, which may prove beneficial for an environment where the teacher is able to focus on instruction. Students from homes with a high education background are likely to bring this to school and thus the education standard in those classes might be higher. Also, in big cities the number of people with low SES and migrant backgrounds might be bigger. With respect to research that found that teachers tend to have lower expectations for students from a low SES and thus are less likely to be disappointed (Speybroeck et

al., 2012), there is an alternative explanation for the potential mastery experience made here. Thus, the community size could serve as an indicator for the school composition in terms of possible support for education, but also level of challenge the teachers have to master. This finding is similar to that noted for ETSE, but given the very small effect size, this explanation is made rather tentatively.

4.3 TEACHERS' SELF-EFFICACY IN CLASSROOM MANAGEMENT (MTSE)

This section contains the results and interpretations of these results of the research question, how teacher-, classroom-, principal- and school characteristics are related to teacher's self-efficacy in classroom management. The explanations will mainly be given by drawing on the effect sizes. It is structured according to the four sub-questions investigating predictors on each level and contains a summary at the end.

4.3.1 BASELINE MODEL

At first, as for the self-efficacy in student engagement and instruction, a baseline model (0) was calculated in order to having a benchmark against which the following models that test the different predictors can be compared. For self-efficacy in classroom management (MTSE), the initial -2log likelihood was 672,867. In total, 13.6% of the initially unexplained variance was attributed to the country level, 2.4% to the school level and 84.0 % to the teacher level. It is notable that there was lower country variance for MTSE than for ETSE or ITSE.

4.3.2 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY TEACHER CHARACTERISTICS?

In the first model, teacher gender (female), educational level, teaching experience, weekly hours spent on teaching, planning lessons and administrative work as well as the extent to which the teacher holds constructivist beliefs about teaching were included in the analysis as potential predictors of MTSE.

The inclusion of teacher characteristics did not explain any further variance on the country nor on the school level. Among teachers, however, 2.4% of the variance could be explained. Compared to the baseline model, the -2log likelihood decreased to 670,525 (**Δ** = -2,342).

When the teacher characteristics are added in the first model (1), all six were found to be significant. Teacher gender (female) was highly significant, but the positive effects were only small and the effect size tiny (**β** = .023, p < .001, ES = .07). This contradicts the hypothesis that male teachers have a higher sense of TSE (Cousins et al., 1996). Klassen & Chiu (2010), however, found that female teachers report lower self-efficacy when facing stress with student behaviour. Here again, results from previous studies are mixed and further investigations into these differences seem worthwhile. However, with respect to previous mixed results, the tiny effect size – although positive – could hint that also in this study neither gender shows a much higher level of MTSE than the other. The educational level effects were also very small, and so is the effect size, but showed a negative relationship with MTSE (**β** = -.011, p < .01, ES = -.03). Comparing this result to the positive association that educational levels shows with ITSE, the opposite relationship could indicate that teachers with a higher educational background feel more confident in their abilities to teach subject content than pedagogically prepared to manage their students' behaviour. This is in line with previous research, which reveals that teachers tend to feel better prepared content wise than pedagogically. Furthermore, in many teacher training courses the focus on pedagogy is often deemed too little (König et al., 2011) and thus teachers might not be able to structure their lessons effectively. All these possible interpretations are made very cautiously due to the very small associations and effect sizes noted here.

Teaching experience showed a positive and stronger association with MTSE (**β** = .005, p < .001, ES = .35), and the effect size is medium, which is in line with the hypothesis and was together with constructivist beliefs about teaching found to have the comparatively largest effect sizes, whereas all other predictors had only very a small effect size. Teaching experience, or 'learning on the job', may provide a useful source for mastery experience, as teachers learn how to deal with difficult students and manage their behaviour, which is supported by Gkolia et al. (2014), who found that teachers with more than 15 years of work experience had

higher levels of MTSE than teachers with less than 15 years of work experience. Mastery experience in general has been found to be a powerful explanation of self-efficacy, which is supported by the size of the effect size here. Additionally, social cognitive theory supports this, as a positive anticipation of successful behaviour management leads to a greater propensity to try out appropriate and effective actions. If these are successful, they are repeated and can lead to further successful actions. This is, again, only partly in line with previous research, where mixed effects have been found for the relationship between TSE and work experience (Cousins et al., 1996; Fackler & Malmberg, 2016; Klassen & Chiu, 2010; Burke & Spero, 2005).

Also, a small positive association with MTSE was found for hours spent on teaching (**β** = .003, $p < .001$, ES = .19) and hours spent doing administrative work (**β** = .005, $p < .001$, ES = .16) and in both cases the effect sizes was small. Both these activities could be seen as supportive of better classroom management. According to Bandura's theory (1978), that states four possible sources of SE (vicarious experience or successful role model behaviour, emotional arousal, verbal persuasion or positive feedback and mastery experience as repeated successful behaviour) they could serve as a source of mastery experience, as explained previously. Hours spent on planning lessons (**β** = -.001, $p < .001$, ES = -.04), in contrast, showed a weak negative relationship with MTSE and the effect size was only tiny. It might be that more effort undertaken in planning lessons can lead to frustration and a lower sense of self-efficacy in classroom management when difficult situations occur, in which the planned lesson is disrupted. Alternatively, teachers may spend more time planning lessons if they have disruptive students (they may have to differentiate work for individuals or prepare different kinds of activities if the students have different needs). However, the effect size is, with respect to the lack of previous research here, too small to draw sound conclusions from. A moderately strong relationship and medium-size effect size was found for constructivist beliefs about teaching that a teacher holds (**β** = .029, $p < .001$, ES = .38) and MTSE. Again, this demonstrates that constructivist beliefs predict different types of TSE and across all three dimensions, the effect size for constructivist beliefs about teaching was the strongest. Thus, this concept seems to be a powerful predictor of TSE.

4.3.3 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY CLASSROOM CHARACTERISTICS?

When classroom characteristics were added to the model (2), all teacher characteristics remained significant. The newly added variables are the class size, classroom climate, low student achievement, the percentage of time spent on teaching and learning, keeping order or administrative tasks, as well as the percentage of students who have a different mother tongue than the language of instruction and a disadvantaged socio-economic background. These two last variables and student achievement provide contextual information of the classroom setting.

Adding classroom predictors led to a decrease of the – 2log likelihood to 663,380 (**Δ** = -7,145). Among teachers, 9.3% of the variance could be explained and 28.6% among schools, but among countries no more variance was explained.

From these newly added variables, class size was positively associated with MTSE and showed a small effect size (**β** = .002, p < .001, ES = .13), as well as a good classroom climate with a strong effect size (the strongest among all classroom predictors) (**β** = .054, p < .001, ES = .99), which is in line with the hypothesis. The positive association and small effect size of low student achievement (**β** = .022, p < .001, ES = .15), in contrast, contradicts the hypothesis that TSE is associated with high student achievement. The strong positive association and large effect size of MTSE with good classroom climate is in line with earlier findings for ETSE and ITSE and also supported by previous research (i.e. Ashton et al., 1983). Furthermore, Rimm-Kaufmann & Sawyer (2004), found that high TSE teachers were more likely to employ stricter classroom management and were more successful in achieving a good classroom atmosphere. Similarly, high TSE teachers were found to successfully apply a diverse set of behaviour management strategies, like strict desists and warnings, but also positive strategies to achieve desired student behaviour (Emmer & Hickman, 1991).

The percentage of time spent on teaching and learning (**β** =.001, n.s., ES = .11) and of time spent on administrative tasks (**β** = .001, n.s., ES = .05) were not significantly associated with MTSE, but had a small, respectively tiny effect size, and the time spent on keeping order was

weakly negatively related to MTSE and showed a tiny effect size (**β** = -.001, p < .001, ES = -.08), in line with the hypothesis based on the literature. Here, the more time has to be spent on keeping order indicates that the students might disrupt lessons and thus the teachers' sense of MTSE drops as they experience a lack of mastery experience and verbal persuasion. Similarly, if teachers are able to spend more time on teaching and learning, they might experience some kind of positive feedback on their classroom management skills. Alternatively, teachers with a high sense of self-efficacy in classroom management organise their lessons so well that they need to spend less time on behaviour management or are able to apply successful strategies (Emmer & Hickman, 1991). The percentage of students having a different mother tongue than the language of instruction was not significantly related to MTSE and was too small to interpret effect size of ES = .01, but the contextual measure of disadvantaged socio-economic family background (**β** = .025, p < .001, ES = .19) was, again, significantly and positively associated with MTSE and the effect size of interpretable size. This positive result is somewhat surprising, as other studies have shown that low SES often occurs with difficult student behaviour (i.e. White, 2010) and poorer achievement (Strand, 2014). An alternative explanation as to why this relationship is positive, could be that teachers have lower expectations for students from low SES and therefore are more easily satisfied and mastery experience is easier to achieve (Speybroeck et al., 2012). This result is, however, in line with those for ETSE and ITSE in this major data set.

4.3.4 To what extent can variation in MTSE be explained by principal characteristics?

In the third model, principal characteristics were added to investigate the possibility of associations between leadership and MTSE. These are: female gender of the principal, age, educational level, work experience and an instructional style of leadership.

Adding principal characteristics led to a decrease of the -2log likelihood (663,369, **Δ** = -11) and the newly added variables contributed to explaining 9.3% variance among teachers, 28.6% among schools, but not to significantly explaining variance among countries.

From the teacher characteristics, all that were significant in the second model remained significant in the third model. The association between classroom characteristics and MTSE, however, did not change either.

From the newly added principal characteristics, only the instructional leadership style was significantly but very weakly related to MTSE and had a tiny effect size (**β** = .003, p < .01, ES = .04), which is in line with the hypothesis that teachers with a principal applying an instruction style of leadership experience higher levels of TSE (Coladarci, 1992; Fackler & Malmberg, 2016). This effect is smaller than that found in models for ETSE or ITSE. This suggests the principal's emphasis on a very active role in this leadership style and modelling behaviour, as well as promoting of a supportive school environment (Caprara et al., 2003; Holzberger et al., 2014) is more important for certain aspects of TSE related to student engagement and instruction rather than behaviour. The other three variables showed effect sizes between ES = -.02 to ES = .00, which are too small to interpret.

4.3.5 TO WHAT EXTENT CAN VARIATION IN MTSE BE EXPLAINED BY SCHOOL CHARACTERISTICS?

In the fourth and last model, school characteristics were added to the analysis: Number of students enrolled at school, school type (private versus public) and community size.

Again, the -2log likelihood decreased slightly to 663,360 (**Δ** = -8). Compared to the previous model, the amount of variance explained on each level did not change and appeared to be non-significant among countries, 28.6% among schools and 9.3% among teachers.

For self-efficacy in classroom management, in the previous models no further variables became significant, but also none became insignificant. Regarding school characteristics, only one, the school type (private), was significant and found to be negatively associated with MTSE but had a tiny effect size (**β** = -.001, p < .01, ES = -.04). With respect to previous research, private schools were expected to have a positive association with MTSE (Fackler & Malmberg, 2016), as those students tend to come from more affluent backgrounds and behaviour problems would tend to be reduced (i.e. Woessmann & West, 2010) and positive and friendly relationships with parents were found to be associated with higher level of TSE, too (Caprara et

al., 2002). Here again, teacher expectations and behaviour management might play a role (Speybroeck et al., 2012): in schools with a more diverse student body, teachers might expect to have to work harder in order to achieve good behaviour compared to those working in private schools, where the relative perception of student misbehaviour might be worse than in public schools. However, due to the tiny effect size, this interpretation is made cautiously. The effect sizes of the other two variables were too small to draw meaningful conclusions (ES = -.01 and ES = .00); especially considering the lack of previous research on them.

4.4 COMPARATIVE MODEL SUMMARIES

The previous sections have presented and discussed the results for each of the domains of teachers' self-efficacy within the scale. However, as in this study the aim is to compare and reflect upon the multidimensionality of the construct, the following section will compare the results across the three dimensions of teacher's self-efficacy by model. With this it is aimed to highlight the similarities and differences in the three areas of self-efficacy in student engagement, instruction and classroom management, according to which the characteristics associated with the teachers, classroom, principals and schools are tested as possible predictors of TSE.

4.4.1 BASELINE MODEL SUMMARY

Most unexplained variance among teachers was found for TSE in classroom management (84.0%), least for TSE in student engagement (67.7%) and the variance proportions were quite far apart. Among schools, in contrast, for all three measures of TSE just 2.4-2.6% of the total variance was attributed to the school level, whereas for countries between 13.6% for MTSE up to 29.8% in ITSE of variance was attributed to level three, the country level. Looking at the difference in the magnitude of total variance attributed to the three different levels, this suggests firstly that working in a different country or individual differences among teachers initially make more of a difference than the actual school in which the teachers work. Secondly, the large differences of unexplained variance between the three dimensions of TSE also support the decision to treat TSE as a multidimensional concept that should be investigated in depth.

4.4.2 1ST MODEL COMPARATIVE SUMMARY

In the first model, teacher gender (female), educational level, teaching experience, weekly hours spent on teaching, planning lessons and administrative work as well as the extent to which the teacher holds constructivist beliefs about teaching were included as potential predictors of TSE. Almost all suggested predictors were significantly related to all aspects of TSE, but varied in terms of the size of the effect sizes; the only exception being the teacher's educational level, which is not significantly related to self-efficacy in student engagement (ETSE).

Teachers' female gender showed a positive relationship with all three aspects of TSE; the strongest though, and medium effect size was found in predicting self-efficacy in instruction. Educational level also had a positive relationship with ITSE, but a negative relationship with MTSE and none with ETSE, and all three effect sizes were tiny. It might also be noteworthy to look at the directionality of the associations: the teachers' educational level showed a positive relationship with ITSE, but a negative relationship with MTSE. The positive relationship with ITSE accords with the view that a higher educational level is a possible source of mastery experience and self-confidence in the teachers' own abilities and understanding of their subject. The negative relationship with MTSE, in contrast, could indicate that teachers who have a higher educational degree and focus more on instruction, as this is what they feel competent in. They feel less efficacious in their abilities to manage their students' behaviour, as pedagogy has often been identified as being underrepresented in teacher training in some countries (König et al., 2011). Further implications of this will be explored in greater detail in the discussion section with particular respect to its potential for teacher training and education policy making.

Furthermore, teaching experience showed the comparatively strongest association with MTSE, but only a weak association with ITSE and ETSE, but all three effect sizes were of medium size and thus this predictor can be considered equally relevant for explaining variation in all three domains of TSE. This pattern of results suggests more years of teaching experience may provide a rich source of learning about behavioural management of students. Again,

mastery experience seems to be a plausible source for growth of TSE during the job experience the teachers gain. This gain seems to be most notable with respect to efficacy in classroom management, rather than in instructional abilities or engaging the students in lessons.

The hours a teacher spends on each - teaching, planning or administrative work in the classroom were similarly positively strongly related to ITSE, ETSE and MTSE and had small effect sizes across all three dimensions. Looking at the direction of the relationship for the hours spent on planning the lessons with respect to MTSE show a negative association, and the effect size was the only one to be tiny. In contrast to the first two dimensions of TSE, where spending much time on preparing lessons could indicate ambitious beliefs in the teachers' instructional abilities, instruction and the opportunities to engage students in the lessons, spending much time on preparing lessons when MTSE is rather low, could indicate a less strong belief in the teachers' abilities to manage their students' behaviour well (perhaps through differentiating work or planning additional activities). Thus, they need to try even harder and prepare better to achieve good student behaviour in the classroom. However, considering the very small effect size, this predictor overall might not be very important for explaining variation in MTSE as compared with ETSE or ITSE.

Finally, the constructivist beliefs a teacher holds about teaching; i.e. that students should be actively involved in knowledge acquisition rather than just being passive recipients, has the strongest association with all three dimensions of TSE and is positive in all three. The effect size, however, is medium for MTSE, but large for ITSE and ETSE. For ITSE, however, the effect size is the strongest across all three dimensions. This shows that the degree of the belief, to which extent teaching quality depends on the teachers' view of students as active participants in knowledge acquisition, seems to be important to explain variation in the teachers' belief in their own abilities to which extent they are able to deliver knowledge. It may be that reciprocal relationships are at work but this cannot be explored in this cross-sectional data set.

As the analysis shows, the effect sizes vary to quite a large extent. For self-efficacy in student engagement, all variables except constructivist beliefs about teaching, which had a medium effect size (.73), had a small effect size (.01 - .34). For self-efficacy in instruction, the effect sizes

were small again (.06 - .39) for all variables except constructivist beliefs about teaching, where the effect size was the largest (1.31) across all 1st models. For self-efficacy in classroom management, the predictors showed varying effect sizes: They were small and positive (.04 - .42) for all variables including constructivist beliefs about teaching, but small and negative for the teacher's educational level (-.03) and the hours a teacher spends on planning lessons (-.04).

For MTSE, only between teachers, additional variance could be explained in the first model, but none on the other two levels. The amount of variance explained was furthermore quite small (2.4%). For ETSE and ITSE, the amount of variance explained among teachers was slightly bigger with 5.7% and 4.2%, respectively. Among schools, however, only for self-efficacy in student engagement could more variance be explained (16.7%). Among countries, then, for ETSE 4.3% and for ITSE 3.7% out of the initially unexplained variance could be accounted for by adding these predictors to the models. As most initially unexplained variance resided among teachers and least among schools, but only relatively little variation among teachers could be explained in the statistical models (1) and – at least for ETSE – most among schools, teacher characteristics apparently contribute to shedding light on differences that are located at the school level, but not so much among individual teachers or countries. This may reflect differences in the way teachers are appointed to or selected by schools in different countries and schools.

4.4.3 2ND MODEL COMPARATIVE SUMMARY

Adding the classroom variables (model 2) did not lead to many changes in the relationship between the teacher related predictors and self-efficacy in instruction, student engagement and classroom management. Compared to the teacher-related variables, the relationship between the classroom predictors and the three dimensions of TSE is more diverse. For predicting TSE in instruction, the class size, student achievement, the percentage of students with a different mother tongue and the percentage of students from disadvantaged backgrounds were found to be significant. Only the classroom climate had a big effect size, whereas all other were either small or medium. For predicting TSE in student engagement, also the classroom climate as well as the percentage of time spent in the classroom on keeping order were identified as significant. Classroom climate showed a positive relationship and had a medium

effect size, while all other predictors showed a small effect size. However, the time spent on keeping order was a negative predictor but had a small effect size. It may function as a form of direct feedback for teachers' ability to actively engage their students in lessons – the more time they need to spend on disciplining students, the worse they might perceive their abilities to actually keep their students interested. For TSE in classroom management, finally, out of the three time-allocation predictors only the percentage of time spent on keeping order in the classroom also has a significant and negative relationship and the effect size was very small, but not the percentage of time spent on administrative tasks or spent on teaching and learning, which had too small effect sizes to interpret them in a meaningful way. Here again, this could be explained by the feedback the teacher is provided by if students misbehave, prompting teachers to spend more time on behaviour management. This, in turn, might act as feedback suggesting own ability was insufficient.

Looking at the effect sizes shows that classroom climate is the strongest predictor across all three dimensions of TSE (.43-.68). Negative effect sizes occurred for two predictors - low student achievement and keeping order in the classroom (-.20 - -.01). For all other predictors, the effect sizes remained relatively small (.03-.43) across all dimensions.

The amount of explained variance changed for all three sub scales comparing model 2 with model 1. For ETSE, among teachers, schools and countries out of the initially unexplained variance significantly more (2.9% among countries, 33.3% among schools and 10.7% among teachers) could now be explained. This is, at the same time, the biggest change happening compared with the second model, where the classroom predictors had been added. For MTSE, among schools, no additional variance could be explained but more among teachers. For ITSE, in contrast, more variance could be explained among teachers and countries. This is somewhat surprising, as many of the classroom predictors appeared to be highly significant. The model fit was again improved.

Furthermore, it is notable that the classroom predictors accounted for a substantial amount of variance at the teacher level. This shows that the teaching context is contributing to the teachers' beliefs in their own abilities of delivering successful lessons.

4.4.4 3RD MODEL COMPARATIVE SUMMARY

Compared to the teacher- and classroom-related predictors (reported for models 1 and 2), only a few principal-related predictors are significantly associated with the three aspects of TSE and revealed only very small effect sizes. The principals' gender (female) and instructional leadership style are both significant for predicting TSE in student engagement and instruction, but for TSE in classroom management, only the principal's gender was significant. For all predictors except the instructional leadership style and ITSE, the effect sizes were > .1. Looking more closely at the two significant predictors identified for ETSE and ITSE, principal female gender is a new finding, which has not yet been noted in previous research. The features of an instructional leadership style, in contrast, are well supported by both theory and previous research. As it reflects the role model behaviour of the principal, it can be suggested that it offers a source of vicarious experience, where teachers are likely to be shaped by the leader's priorities and through practical guidelines or tools that they can use in their lessons. Furthermore, the academic focus and goal-setting aspect of this leadership style may be a source of support and guidance for the teachers' work and belief in their ability to make a difference to student outcomes. With respect to MTSE, where only the instructional leadership style was identified as significant, it suggests that the support the principal offers in teaching is an important aspect. It is a potential source of both mastery experience was well as vicarious experience, as the principal offers strategies for the teachers to imitate to successfully manage students and allow them to anticipate success in the classroom. Interestingly, neither the principals' age, work experience nor educational level were significantly related to any of the three dimensions of teacher self-efficacy or had an interpretable effect size. With respect to age and work experience, TSE theory suggests that older and more experienced principals might offer a richer source of vicarious experience with respect to their role model behaviour and could also more strongly engage in, for instance, goal setting. However, other research on school leadership generally finds that these characteristics of school leadership are not related to student academic outcomes and the present findings move on to suggest they also have little influence on aspects of TSE.

The effect sizes identified for all principal characteristics and their associations with the three dimensions of TSE are small (.04-.12). The largest effect size was identified for instructional leadership and ITSE, whereas the effect sizes for female gender and instructional leadership and self-efficacy in instruction, student engagement and classroom management were very small.

However, compared to the teacher- and classroom characteristics, the variation in principal characteristics does not seem to make the same difference for TSE. For ETSE, on all three levels more variance could be explained: 4.3% among countries, compared to 3.7% for ITSE and none for MTSE. For ETSE among schools 33.3% of the initially unexplained variance could be explained, 28.6% for MTSE but no more for ITSE. Then, for self-efficacy in student engagement 10.7% of the variance could be explained among teachers, whereas for self-efficacy on instruction only 6.3% and for MTSE 9.3% of the variance could be explained. Here, the same pattern as in the previous models remains; that in total most of the variation among schools in TSE could be shed light on, but differences between countries and teachers were little affected by the set of principal predictors in this study. Finally, the model fit (model 3) improved for all three dimensions of TSE compared to the second model.

4.4.5 4TH MODEL COMPARATIVE SUMMARY

Adding school-related predictors (school enrolment, public vs. private school and community size) to the analysis showed different results for the three sub-scales of TSE: Student enrolment has no significant association with either dimension of TSE, whereas teaching at a private school is found to be weakly negatively, but significantly related to MTSE and ETSE, but not ITSE. The community size also shows a significant association with two aspects, ITSE and ETSE, but all effect sizes were rather tiny. The effect sizes of the school predictors in the last model were, again, small and ranged from -.01- -.11. The relatively largest effect size (-.11) was found for ETSE in relation to teaching at a private school, whereas the smallest effect size (-.01) appeared for MTSE in relation to the size of the community, in which the school is located.

Adding those variables did not change the relationship between the three sub-scales of TSE and the predictors of the other models compared to the previous model. Accordingly, on the teacher level for ETSE 10.7%, for ITSE 6.3% and for MTSE 9.3% variance could be explained. Among schools, no additional variance could be explained for ITSE, but 33.3% for ETSE and 28.6% for MTSE. Finally, among countries, 4.3% of the variance for self-efficacy in student engagement, 3.7% for self-efficacy in instruction, but no more for self-efficacy in classroom management could be explained. In other words, most of the variance between countries in the three measures of TSE was little affected by these school characteristics. For the teacher level, school characteristics led to modest additional statistical explanation. For school level variance, however, this is very different and the full model contributed to explaining a majority of the initially unexplained variance for all dimensions except ITSE. Small differences among countries, finally, could be explained by the full model; except for MTSE, where the initial amount of unexplained variance remained. The model fit again was improved for all three sub-scales.

4.5 SUMMARY

The results of the multilevel analysis discussed in this chapter supports the decision to treat TSE as a multi-dimensional concept, as the same predictors show different patterns of relationships with each of the three aspects of teacher self-efficacy considered. In particular, the statistical significance and also the magnitude of the effect in terms of the effect sizes vary – sometimes by a larger amount – for various predictors across the three dimensions.

To sum up the results of these analyses, it can be said that the testing of the characteristics that predict tripartite teacher self-efficacy revealed mainly five findings:

1. Each domain-specific TSE is associated with teacher-, classroom-, principal- and school characteristics in a different way. Although there are similarities, some of the predictors (i.e. the different associations of time allocation in the classroom in model 2) show differential effects that offer implications for further research in the future. The domain specific literature on TSE, especially, helped further explain differences between the three domains.

2. The effect sizes all remained rather small, with only a few exceptions (i.e. the measure of constructivist beliefs about teaching, and that of classroom climate) across all models. They were particularly small for school and principal characteristics.

3. Proportionally, most variance could be explained among schools. Nonetheless, this should be interpreted cautiously in the light of the null baseline models, where only a small amount of total variance resided among schools. Accordingly, the individual differences among teachers seem to most strongly contribute to differences in the perception of self-efficacy, followed by the country, in which the teacher works.

4. Most differences in the association between the dimensions of TSE and the suggested predictors seem to occur among classroom characteristics. One particular characteristic is that within the classroom, however, most human interaction between teachers and students happens and the direct environment in which teaching takes places is developed there. These different associations across the three dimensions show that teachers' perception of their own efficacy in different domains is related to classroom features in different ways and are not necessarily common predictors.

5. The explained variance among schools on the one hand and teachers and countries on the other hand seems to behave rather differently for the dimensions of TSE: ITSE seems to have a different pattern than that found for ETSE and MTSE which are more similar in terms of the findings. The models for ITSE did not reveal many changes between the second and fourth model, as the amount of explained variance remained stable among schools and teachers and only changed slightly among countries. For ETSE and MTSE, in contrast, the amount of explained variance differed markedly across the models.

These results, again, support the view of TSE as a multi-dimensional concept, as the same predictors show differently strong relationships with each of the sub-concepts and also the magnitude of the effect varies across the dimensions. The different amount of variance that could be explained on each level across the ITSE, MTSE and ETSE also support the concept of a multi-dimensional construct.

The next chapters contain some additional methodology as well as the results and interpretation for research question two. After that, a final discussion and conclusions chapter will provide further consideration of explanations and reflections with respect to theory and previous research and will bring the two research questions together by examining implications of the findings for educational decision making and possible practice.

CHAPTER 5 COUNTRY-SPECIFIC CHARACTERISTICS OF SCHOOL SYSTEMS AND SOCIETIES

In international organisational research, it was possible to account for some of the variation in self-reported employee behaviour with respect to various measures of cultural values and beliefs, as has been done by House et al. (2004), where the norms and values of 62 societies have been mapped, which has been the most comprehensive study on this topic so far. Hence it seems worthwhile to explore whether such associations might be evident in relation to teachers by investigating, how teachers in different cultural contexts respond to the environment in which they work. Schools, as such, might not be international organisations, but they prepare students for a globalising world. Furthermore, there are a number of teacher characteristics, like teacher's self-efficacy, that are deemed important internationally, regardless of the norms and values that shape the environment in which the (school-) culture is developed. No research so far has investigated whether TSE may be shaped by country contexts in terms of different cultural values and beliefs. In order to address the lack of research, in the following predictors of teachers' self-efficacy will be compared across three different country clusters. Those clusters comprise three Anglo-Saxon- (England, USA, Australia), four Nordic- (Denmark, Finland, Norway, Sweden) and four East- and South-East Asian (Japan, Singapore, South Korea and Malaysia) countries.

Thus, this chapter has three main aims: Firstly, it seeks to characterise the school systems that form the three different country clusters by identifying commonalities between them. The framework used here was suggested by Bulle (2011), who developed a system to group countries according to their academic vs progressive orientation of the school system and the degree of differentiation of the curriculum during secondary education. This will provide a theoretical reference for grouping the countries into clusters. Secondly, it seeks to explore literature concerning possible core values that may shape the societies' understanding of the relationships across hierarchies, which in the school context, defines the relationship between principals and teachers or teachers and pupils, respectively. I will mainly refer to Hofstede's (1986, 2010, 2011) analysis of leadership values here and complement those with evidence from the GLOBE study where necessary. Thirdly, it seeks to provide an overview of how country cluster's specific attitudes towards teaching are displayed and applied in each cluster.

5.1 CATEGORISATION OF EDUCATION SYSTEMS

Bulle (2009, in: Bulle, 2011)[9] provides a comprehensive framework to cluster different education systems according to two main institutional characteristics: firstly, their curricular structures during compulsory/lower secondary education, which can be undifferentiated, mixed or differentiated and secondly, their educational objectives, which can have an academic focus or a progressive focus. In these systems the focus is either on academic objectives like high scientific standards, which characterises the academic type, or acquiring competencies like the application of concepts, which characterises the progressive type. In summary it is argued that the academic type lays emphasize on the structured learning and understanding of subjects in a scientific way; i.e. the understanding of mathematical or physical concepts or the grammar of a foreign language and hypothetical-deductive abilities rather than only training the application of it. Pupils' more general skills, in contrast, are the focus of those school systems that are typical for the progressive type. These are seen to apply empirical learning, competence-based courses and follow more constructivist teaching approaches like a focus on group rather than whole class or individual activities. Theoretical teaching, in contrast to the academic approach, only plays a subordinate role, but practice, immersion and inductive skills are promoted according to Bulle's framework.

Furthermore, the degree of differentiation of the school system is another important feature. Undifferentiated courses are characterised by a long time, i.e. the end of compulsory education, the pupils learn together and share a common core curriculum and make little use of academic selection (e.g. comprehensive schools in the UK) and only specialise at a late stage of their secondary education. By contrast, the differentiated type uses academic selection to sort pupils at a younger age. For example, pupils are encouraged to choose between different tracks (i.e. low, medium and high academic track) in Germany. Finally, the mixed type in some way bridges the two more extreme approaches, as it is characterised by allowing for variations of i.e. a strict core curriculum that all schools in one country or district must adhere to

[9] In this paper, Bulle summarises the analysis she has provided in her book: *Bulle, N. 2009. L'école et son double. Essai sur l'évolution pédagogique en France. Paris: Hermann.*, which is only available in French. Thus, the article will be referenced here instead.

by allowing for local variations that, for example, take into consideration the students' aspirations in key subjects such as sciences and mother tongue, when offering further secondary education (e.g as applied in the Netherlands).

Applying this categorisation to the available TALIS countries, brings up three clusters: The Anglo-Saxon cluster[10] with the USA, England and Australia, the Nordic cluster with Finland, Sweden, Norway and Denmark and finally the East- and South-East Asian cluster with South Korea, Japan, Singapore and Malaysia. Here, the Anglo-Saxon cluster can be seen as an example for a broadly mixed system with a progressive trend, as on the one hand the focus is on academic courses early on in secondary education, but on the other hand pupils attend courses together as a group until the last few years of their secondary education and selection is not a major feature. Furthermore, the academic objective lies on achieving competencies rather than on rigorous academic schooling. More academic schooling starts with upper secondary education. In England, for example, entry to upper secondary education (A-Level) is based on the national exam (GCSE) grades[11] (Smithers & Robinson, 2010) and is also voluntary for students (they must remain in education or training till age 19 but this is often vocational not academic education). Similarly, in the USA, upper secondary education is voluntary, too, and serves as a preparation for university. In Australia, in contrast, upper secondary education is compulsory. In both countries equally, progression into upper secondary education is mainly by age and teacher recommendation (ibid.).

The Nordic cluster, in contrast, is characterised as an undifferentiated cluster, but with a more progressive orientation. In the Nordic countries, a core curriculum is taught mostly until the age of 16, after which pupils can decide whether they want to continue their academic education or move on to a vocational track. Furthermore, the trend the Scandinavian countries follow in education is progressive. Thus, the focus is on the development of competencies,

[10] In the following, I will use the terms 'cluster' and 'countries' interchangeably to avoid repeating each word too often. Thus, 'Anglo-Saxon countries' and 'Anglo-Saxon cluster', for example, refer to the same analytical entity, as the countries constituting the Anglo-Saxon cluster.

[11] Apart from academic results i.e. the parents' financial possibilities play a role in the choice of upper secondary schools due to fees that apply at certain schools. However, this will not be discussed further here, as the focus is on the academic, not the social requirements for upper secondary education

empirical learning and inductive approaches, which means the learning is focused on making inferences from specific examples to general rules. However, although the Nordic countries tend to be undifferentiated officially, their school policies allow for individualisation of the curriculum by i.e. offering special classes according to the pupils' interests. Thus, despite the egalitarian approach to education it could be argued that in fact 'hidden streaming' is taking place by allowing the pupils to cluster into groups according to their interest or ability in certain subjects, for example. math or history. Another feature of this cluster is the selection into upper secondary education that takes place after lower secondary education, which marks the end of compulsory education in these countries (OECD, 2004). In Finland, Norway and Sweden, for example, the key criterion is the recommendation of the feeder school, while Denmark considers the recommendation of the feeder school or the students' academic performance to be equally important. Thus, although the system itself is quite progressive and does not follow the academic trend, which advocates for performance orientation, performance in lower secondary education is considered to be important and plays a key role in shaping the students' future career paths.

Finally, the East- and South-East Asian cluster is classified as a mixed cluster with an academic trend. It is mixed for the reason that, despite a strict core curriculum, there is room for schools to offer extra courses according to the pupils' ability and aspirations, select pupils according to specific criteria and allow for private schools to differ from the core curriculum and complement public education. In South Korea, for example, in addition to prior school performance, the recommendation of the lower secondary school is taken into account and residence in a particular area. In Malaysia, in contrast, upper secondary education is granted for everyone, but prior results determine the type of schools the pupils are eligible for (Mokshein et al., 2011) and Japan requires its students to complete entrance exams for their chosen upper secondary school (Abumiya, 2012). Thus, the education received in lower secondary schools also plays an important role in these school systems as it in various ways determines the future academic career of the pupils and reflects the strong emphasis on academic education and selection.

It has to be noted, however, that Bulle's characterisation of school systems and the country clustering emerging from it is only one of several possible solutions. House et al. (2004), in

their world-wide study investigating 62 societies among others, chose to cluster countries according to their historic, cultural or language proximity. For the Anglo-Saxon cluster, language also seems to be a unifying element, as is cultural or geographic proximity for the Nordic cluster. Given the clear focus of the investigation of schools systems and differences in the perceptions of teachers' self-efficacy in relation to predictors concerning the teachers themselves, the classrooms they are teaching in, the principals who are responsible for school management and school characteristics more broadly defining the context in which the teaching takes place, Bulle's approach, i.e letting characteristics of the school systems guide the clustering of countries, was chosen as the most appropriate for this research study. Moreover, elements of cultural or geographical proximity seem to be reflected in the similarities of the school systems, which, in this case, provides this clustering approach with the stronger and richer explanatory power.

5.2 SUMMARY

As can be seen from the categorisation above, there are distinct features of each of the school systems that make them similar to each other, but different as well. For the Nordic countries, this includes the length of time pupils learn together before splitting into different tracks or the strong focus on deductive and abstract skills, and among the East- and South-East Asian countries, selection. Nonetheless, it has to said that these categories are only broad heuristic devices reflecting a possible 'ideal-typology'. There are irregularities within them, as it should be noted that even in Nordic countries there are private schools, which differ from the public system. Similarly, in the UK there are private schools and a small number of selective grammar schools, while social selection may also occur reflecting differences in social geography. However, these irregularities are not considered strong enough to change the overall classifications of education systems into the three broad groups that will form the basis for further analyses of research on variations in TSE.

Another feature is that lower secondary education, in which the teachers were investigated in the TALIS study work, in most countries, leads on to higher secondary education. Especially in those countries where there are entry requirements into higher secondary tracks, it can be assumed that special emphasis is put on the abilities and convictions with which the teachers

teach. Thus, differences and similarities in teacher's self-efficacy and predictors of it might be particularly interesting to investigate.

5.3 COUNTRY VALUES AND LEADERSHIP

To further distinguish the above-introduced country clusters from each other, overarching values that predominate in the clusters will be further explored here. This idea is drawn from Sternberg (2004), who describes cultures as a set of values, beliefs and attitudes that are shared by a group of people. Thus, they provide a lens through which the world is perceived, and decisions are made, so helping us to distinguish one society from another (Hofstede, 2011).

It is demonstrated by, for instance, Hofstede (1986) that cultural values are important for the understanding of the functioning of classrooms and educational settings, as they form the environment in which the school system and thus teacher and pupils' behavioural patterns are developed. Evidence about such country values will be taken from Hofstedes' research, as he initially introduced the concept of country- and leadership vales and from the GLOBE study, which also investigated cultural beliefs and practices in 62 countries with respect to nine different dimensions (House, Hanges, Javidan, Dorman, & Gupta, 2004). Four of these sets of values are deemed to be particularly important for the school context: power distance, gender egalitarianism (or femininity vs masculinity), collectivism vs individualism and performance orientation. Those will be introduced in the following sub-sections, with particular attention paid to identifying which values are more likely to prevail in each of the three country clusters. The table below gives an overview of the means of each country for these four dimensions within the cluster, and also presents means for each country in the clusters (c.f. House et al., 2004). For gender egalitarianism it needs to be pointed out that a higher value means less gender inequality and for individualism/collectivism, a higher value means a higher degree of collectivism and a lower accepted degree of individualism in the respective clusters.

Table 21 Means of the country values by cluster and country

	Nordic				Anglo-Saxon			ESEA			
Performance Orientation	5.75				5.98			5.55			
Gender Egalitarianism	4.82				5.08			4.21			
Power Distance	2.55				2.81			2.85			
Indvidualism/Collectivism	4.08				4.58			4.33			
	SWE	DEN	FIN	NOR	USA	ENG	AUS	SIN	JAP	SKO	MAL
Performance Orientation	5.80	5.61	6.11	5.49	6.14	5.90	5.89	5.72	5.17	5.25	6.04
Gender Egalitarianism	5.15	5.08	4.24	4.99	5.06	5.17	5.02	4.51	4.33	4.22	3.78
Power Distance	2.70	2.76	2.19	2.45	2.85	2.80	2.78	3.04	2.86	2.55	2.94
Indvidualism/Collectivism	3.94	4.19	4.11	4.55	4.17	4.31	5.27	4.55	3.99	3.90	4.87

5.3.1 POWER DISTANCE

Power Distance is defined as the degree to which members of a society believe that power should be distributed equally across hierarchies. A high degree of power distance thus means that the accepted differences between hierarchies are big, whereas a low degree of power distance is an indicator for flat hierarchies where the accepted differences are rather small (House et al., 2004). Countries with small power distance still accept the use of power, but only with legitimate reasons and hierarchies are only established when needed but are not assumed as a given. This also includes that subordinates (i.e. teachers in schools in relation to their principals) expect to be part of decision-making processes. When it comes to educational settings, schools in societies with small power distances tend to use student-centred teaching approaches (Hofstede, 2011). In contrast, societies that accept a large power distance between its members treat hierarchy with great respect. This is also visible in the relationship between bosses and their teams, who tend to expect to be given instructions, but do not expect to participate in the decision-making processes themselves. In the classroom, teaching is rather teacher-centred, and pupils are expected to follow the pace and the rules that the teacher establishes. Thus, with respect to the classroom and the school environment, this understanding shapes the relationship between the teachers and their pupils, or the teachers and principals respectively (ibid.). These differences also occur in the country clusters. In Anglo-Saxon countries, power distance tends to be relatively small, meaning that hierarchies tend to be rather flat and that the power between employees and their superiors should be shared more equally. In East and South-East Asian countries, the accepted degree of power distance is comparatively small again, but higher than in the Anglo-Saxon countries. Thus, with respect to principals and teachers, the principal's hierarchical position is greatly

respected. Finally, in the Nordic countries, the degree of accepted power distance is the smallest among all three country clusters.

With respect to teacher-student relations, there are also differences between cultures with small and large degrees of power distance. In cultures with a large power distance, as in the ESEA countries, the teacher is highly respected both in- and outside the classroom, is expected to perform active leadership in the classroom by, for example, initiating communication and conducting mainly teacher-centred lessons and in conflict situations the teacher is more typically supported by the parents. In cultures with a smaller degree of accepted power distance, which are here represented by the Nordic (m_{PD} = 2.55) and to some degree the Anglo-Saxon (m_{PD} = 2.81) countries, accepted models of teacher-student interaction are different. Overall it can be said that the lessons are – compared to the ESEA (m_{PD} = 2.85) countries – expected to be more student centred. Thus, the students are allowed to question their teachers, communication can be initiated by both sides, teachers and students, students are expected to act more independently, and in conflict situations, parents might often support the students and not necessarily the teachers, in contrast to the case in countries with high degrees of power distance (Hofstede, 1986).

5.3.2 GENDER EGALITARIANISM

Secondly, *Gender Egalitarianism*[12] describes the degree to which gender differences in societies are minimised. Accordingly, a high degree of gender egalitarianism stands for a low acceptance of gender differences. These gender differences mainly concern 'stereotypical' views on which attributes, behaviours and character traits are regarded as being 'male' or 'female' (House et al., 2004). Examples for so called male traits are seen to be strength and dominance, examples for so called female traits would include gentleness or empathy. In more egalitarian societies, there are more women in leadership positions and in the workforce and family, work tends to be more equally distributed between men and women. In less egalitarian societies, in contrast, which can rather be described as more 'masculine', women

[12] Hofstede's corresponding categories are ‚femininity' vs ‚masculinity', with 'feminine societies' having little differences between genders and 'masculine societies' having large differences between gender roles

are not very well represented in the workforce or leadership positions and are not expected to be assertive and ambitious, which is seen as only appropriate as a characteristic of men. In the Anglo-Saxon countries (m_{GE} = 4.45), gender egalitarianism is generally less highly supported, as it is in the Nordic countries (m_{GE} = 4.82). In the ESEA countries (m_{GE} = 4.21), gender egalitarianism is supported to a lesser degree than in the other two country clusters. Given the communist history of these countries, where differences between men and women are traditionally less supported, this result is somewhat surprising. However, somewhat different results are found by Hofstede et al. (2010), where masculinity is found to be high in Japan, but rather low in other East and South-East Asian countries. In Anglo-Saxon countries, masculinity was only moderately high and rather low in the Nordic countries.

These tendencies also seem to prevail in the three country clusters, which are part of this study. In societies, which show more acceptance of segregated gender roles, there is a high recognition of ambitious teachers and competition in the classroom is supported through reward systems and open praise of good students. In contrast, in societies where gender roles are partly interchangeable, there is less open competition in the classroom and teachers may focus more equally on all students rather than just the most able high achievers, avoid open praise of student attainment and are appreciated for being friendly and approachable and concerned for student well-being (Hofstede, 1986).

5.3.3 COLLECTIVISM AND INDIVIDUALISM

Thirdly, *collectivism and Individualism* is another feature that may be relevant. Societies that put high value on collectivism are seen to appreciate loyalty, pride and cohesiveness at the workplace or within families. Belonging to a group (i.e. the larger family, colleagues at work) tends to be crucial for their self-esteem and identity. Their communication style tends to be indirect, their identity is formed in relation to the group and social interactions tend to be long term. Individualistic societies, in contrast, value individual goals above group goals, communicate in a rather direct way, think rather rationally than empathetically and people tend to engage more often with each other and establish short-term relationships. People in individualist societies take most interest in the immediate family or close friends rather than broader social grouping (Hofstede 1986, Hofstede 2011). In the ESEA countries, institutional collectivism is supported to a fairly high degree. Some more emphasis on individualism is

found in the Anglo-Saxon countries, but by far the largest support for individualistic attitudes is found in the Nordic countries, where the support for collectivistic values is found to be low compared to the other countries.

There are differences in the classroom behaviour between societies which support collectivism, as can be seen for the ESEA (m_{IC} = 5.02) countries, and societies that rather support individualistic values like the Anglo-Saxon (m_{IC} = 4.19) and Nordic (m_{IC} = 4.08) countries. In classrooms in collectivistic societies, it is suggested that it is more accepted that students form cohesive sub-groups that are based on ethnicity or religious affiliation, a good classroom climate is considered to be important and students only speak when called upon by the teacher (Hofstede, 1986). Furthermore, education is seen to have high social prestige, which is achieved by attending 'the right schools' or gaining certificates. In more individualist societies, in contrast, sub-groups are formed in a rather pragmatic way (i.e. based on availability of students or task-based), classroom climate can be open and confrontational, and students are more willing to speak upon general invitation 'to the class'. Education, as such, is rather seen as a means to gain skills and qualifications and self-respect, but not so much belonging to a certain group. Thus, certificates are considered less important than in collectivist societies and a larger emphasis is put on competence (according to Hofstede, 1986).

5.3.4 Performance Orientation

Finally, *Performance Orientation* is considered to be relevant to this research as a means of striving for excellence and improvement in societies. Societies that lay emphasis on performance orientation are thought to set demanding targets, value individual achievement, believe that success can be achieved through hard work and see schooling as a means for success. Societies and organisations, in contrast, that have a low appreciation for performance orientation, value loyalty over success, emphasize harmony over merit and demand respect for seniors. Moreover, they regard the choice of an appropriate school as crucial for success in later life, value sympathy and build on tradition and belonging to networks. Compared to the other country values, performance orientation had the greatest support in all three country clusters. In the Anglo-Saxon (m_{PO} = 5.98) countries the support was comparatively higher than in the Nordic countries (m_{PO} = 5.75), and in the ESEA countries (m_{PO} = 5.55), which scored the lowest among all three country clusters (House et al., 2004).

5.3.5 SUMMARY

The section above aimed at describing four dimensions of country values and to which extent they are supported in each of the country clusters: power distance, gender egalitarianism (or femininity vs masculinity), collectivism vs individualism and performance orientation. It seems to be noteworthy that this not a clear dichotomous categorisation, but countries tend to lean towards these attitudes. Thus, characteristics of one or the other category can be found in a country cluster, which largely falls in one domain, without contradicting this.

5.4 TEACHING STYLES AND CULTURE

Apart from different leadership styles and values that are promoted to a different extent in the different countries, also teaching styles in the classroom vary in different cultural settings and follow different pedagogical and societal aims. This section will provide an overview of differences in classroom practices and teaching patterns in Anglo-Saxon, Nordic and East- and South-East Asian countries to further contextualise the environment, in which teaching and learning take place. With this it is hoped to shed further light on differences that occur between predictors of TSE in the three country clusters. It furthermore aims at briefly describing the aims and underlying philosophies that shape the education systems in each country. Those 'aims' are usually not part the curriculum, but are stated as overarching structures, in which education and schooling take place. Thus, they are, for example, found in the classroom structure, the relationship between teachers and students and how knowledge is approached (c.f. academic vs. progressive education according to Bulle, 2011), which are meant to teach accepted social behaviour and prepare students for society. It is not the aim of this section, though, to give a comprehensive overview of these features, but rather to further contextualise the findings and provide a framework for the interpretation of them.

5.4.1 TEACHING AND EDUCATION IN ANGLO-SAXON COUNTRIES

As it is the case with the other country clusters, there are quite a few differences between the Anglo-Saxon countries in the cluster, but certain basic values also seem evident that may unite them in such a way so that they share a common set of beliefs, which appears to shape their education systems. It can be argued that for the contemporary Anglo-Saxon education systems, John Dewey and his idea about 'democratic education' are central (here and in the

following: Dewey, 1899, 1902, 1916). Dewy suggested that, in addition to knowledge transmission, schools should prepare pupils to be future citizens and include social skills and community life in the curriculum. The traditional or academic strand was characterised by early differentiation between academic schooling in preparation for university studies and vocational training in preparation for the labour market. Dewy criticised that this way of schooling would on the one hand create a gap between those preparing for academic studies and those, whose track would necessarily lead to manual work. Apart from expanding the duties of schools to promote non-academic goals, Dewy fostered the idea of 'education for all', regardless of social background and parental education, which he saw best promoted in a system where also vocational education would be offered alongside academic education. With this it was hoped that not only the gap between rich and poor would become smaller, but that also, people, who had the potential to attend academic education but would not have considered it otherwise would be attracted to it.

Therefore, he suggested an education strategy that should counterbalance these shortcomings by using teaching methods and topics that not only deliver academic content knowledge, but also require the students to engage in activities that involve problem solving, critical thinking and cooperative learning projects to develop not only content knowledge, but also, for example, democratic thinking or social responsibility.

These ideas have wider implications for the setup of the education systems, the way in which teaching, and learning are approached in those countries and the role of the teacher in the system.

In America, it has been suggested that there is an emphasis on learning tools to apply knowledge (Ravitch, 2008), which means that not only theoretical knowledge, but also the understanding of the concepts that are needed to apply the knowledge, are taught at school. This is put in practice, for example, by applying strategies like group teaching, which embraces the ideas of a democratic exchange of ideas among pupils, weaker pupils get the chance to learn from stronger pupils and communication skills are trained by engaging in the negotiation of ideas. Thus, the role of the teacher often is one of the 'supervisor' of the class. The second teaching strategy that is frequently applied are teacher-led sessions, where whole-class instruction takes place. The idea behind this is that all students should have an equal

chance to learn from the teacher's expertise and no pupil should be left behind (Norris, 2004). A study by Mujis et al. (2018) showed that American teachers are strong in using assessment and evaluation as tools for teaching that provide feedback on the students' current abilities and structuring lessons by using a variety of teaching strategies. This includes checking individual student's understanding during seat work or consolidating new material by solving exemplary tasks together and providing real-life scenarios that provide a context for the lesson content. A frequently used tool was group teaching that led to interactive lessons that fostered student communication on the one hand, but also provided opportunity for differentiation as faster groups could move on more quickly when they had finished one task. Another small-scale study by Lindorff & Sammons (2018) again showed that American teachers can use both, teacher- and student-centred approaches, but found that the use of it varies across classrooms, but not always within it. They could also show that the success of especially the student-centred approaches was linked to the classroom management skills of the teacher, as stricter teachers were more successful in providing feedback and ensuring learning success even in more individualised settings. Thus, the approach formulated by Dewy that the classroom should provide a space for content and social learning, was visible in the American classrooms, although also traditional teaching approaches are used.

In England, however, the education system looks somewhat different from the one in the US, despite it also being influenced by Dewy and pragmatism. Most schools there are publicly funded and non-selective and offer different foci and opportunities for the students, as, for example, job-oriented practical courses alongside the usual curriculum. Furthermore, there are a few state selective grammar schools that have a strong academic focus. Alongside those, there are some highly prestigious private schools. Entry into these schools is mostly selective and entry exams are required (Eurydice, 2010). In a study of secondary schools it was suggested the most effective way of teaching in England involves teachers being clear about their goals, have a 'toolbox' of various teaching strategies at hand so that they can adapt their way of teaching to both the needs of the students and the requirements of the subject area and teach students meta-cognitive strategies so they can reflect upon their own learning process. They furthermore involve the students frequently in the lessons to monitor their learning process, provide an opportunity for guided practice and quick feedback and make link to other subjects where content overlaps with other subjects. Hence, successful teaching in the

UK can be sees as a combination of progressive constructivist student-centred and traditional teacher-centred approaches towards teaching (Ko & Sammons, 2016).

The third country in the cluster is Australia. Australian teachers interpret their role with mainly three different foci according to Boulton-Lewis et al. (2001): first of all, a student focus, where they aim at developing the students' cognitive, behavioural and affective skills by providing the opportunity to engage in a variety of different activities. Secondly, the focus is on competencies and knowledge, which has in focus the learning of a certain type of knowledge and its application in praxis. It covers both, taking in new information, for instance, mathematics, and at the same time, to apply it to certain praxis related tasks. Teachers tend to use both, small group and whole-class teaching methods and switch between them smoothly. Furthermore, Australian teachers make an effort to use active teaching methods that engage students in the lessons and stimulate conversations between students. However, Australian teachers, although reportedly having a lot of freedom to design their lessons, feel under pressure from official authorities and superiors to fulfil duties over and above classroom teaching, lesson planning or assessment: they are also expected to provide extra activities for students to further engage in topics. Perceived official pressure was reportedly higher than pressure from parents (Aldridge et al., 1999). When it comes to teacher respect and classroom discipline to provide a supportive learning environment, Australian teachers report a rather low level of unconditional respect they receive from students; especially in secondary education. Nonetheless, teachers aim to teach their students respect and to take responsibility for their own behaviour (Lewis et al., 2005). Here again, the focus is on the use of communicative student-focused methods rather than on a teacher-led setting (ibid.). As with the USA, Australia has a strong emphasis on John Dewy and his idea of democratic education (Appleyby & Limm, 2018). School is seen as a main means to inculcate these values and the current academic debate evolves around the question of how and to what extent Dewy's ideas can be brought back in a culture that largely emphasizes a technocratic approach towards schooling and education (Webster, 2017).

5.4.2 SUMMARY

As can be seen from the description of the three Anglo-Saxon school systems above, they all seem to have been influenced by the idea of Dewy's democratic education and a potential

focus on constructivist teaching methods, but they are nonetheless very different from each other. In England, most students attend publicly funded schools and only a minority is educated at highly prestigious private institutions. It allows for a broad variety of state schools and thus offers a lot of choice. In America, the origin of the democratic education approach, studies have found that there is indeed a focus on student-centred constructivist teaching methods, but also traditional teaching style is used. Finally, in Australia, a similar situation as in America has been observed: Teachers see themselves as responsible for facilitating learning and the acquisition of competencies but make use of a toolbox of various teaching strategies that they deem appropriate for the respective goal. What all three have in common, however, is that teachers have a high social status and are awarded respect in public (Hargreaves, 2009), but nonetheless often face difficulties to discipline students and enforce this respect in the classroom. It can be speculated that this comes with a rather progressive student-centred approach at schools, where students are given more power than in the traditional teacher-centred setting.

5.4.3 Teaching and education in East- and South-East Asian countries

Although the East- and South-East Asian countries in this cluster are quite different from each other, they share being grounded in Confucianism as a common denominator, which helps to explain basic values and traditions that shaped the education system, too (Pang, 2006). Confucianism, in essence aims at living in harmony with both, people and nature, around oneself. Key to this are obedience to the elder (i.e. teachers or parents), striving for excellence through hard work to be a good 'son and citizen' and the development of a strong and virtuous character (Gardner, 2014). Hence, at school, great respect for teachers is demanded. Furthermore, a huge emphasize is given to effort, strenuous learning and memorising as well as emulating the teachers and getting as close to their level of expertise as possible. Hereby, the teachers have a crucial role as the school is seen a place where both, the forming of a strong character and the transmission of desired values and the striving for academic excellence take place. Moreover, in the age of globalisation, Asian countries aim at being recognised internationally and are undertaking huge efforts to support education both nationally and privately. Nationally, this means that attention is being paid to providing schools and teachers with the

resources they need to deliver high-quality education that results in high-staking exams. Privately, parents are making great financial investments and personal sacrifices to make sure their children are provided with as-good-as-possible education that enables them access to prestigious universities and jobs (Pang, 2011). Nonetheless there is much concern about the high levels of pressure placed on students to succeed and the impact on mental health (Zhao, Selman & Haste, 2015).

To put this in practice, in East Asian countries, the predominantly teaching method is 'whole-class teaching' (Stevenson & Lee, 1995). In Japan, for example, the lessons are delivered by the teachers, who see themselves as 'informed guide' (ibid. p. 154), to the whole class. Children tend to mainly spend time working, listening to or watching their teacher as a class instead of doing group work. Nonetheless, there is active discussion and knowledge exchange between the teacher, the students going on, as the teachers provide critical feedback on the students' answers and those answers are brought to discussion with the whole class, too. With this, the teachers make sure the students develop a deeper understanding of the topic. Another type of instruction the teachers use are individual worksheets or small group teaching, where the teacher moves through the classroom and provides individual feedback to single students or student groups to check on the students' abilities and being able to provide for the students' individual needs by, for example, by presenting material in different ways so that fast and slow learners can be accommodated in one classroom. Predominant characteristics of the classroom in, for instance, Shanghai is that the teacher applies a variety of techniques to address the students and the topic in the most suitable way, using precise language to make sure the topic is conveyed unambiguously, and a logical structure of the lesson is presented (Lim, 2007). Furthermore, classroom discipline and teacher respect are taken very seriously. A good student-teacher relationship is maintained and teachers among themselves emphasize collegiality (ibid). Another feature of ESEA classrooms is the way in which the teacher is being treated as a respectable person and has a high social status, as the students are raised that way (Sato & McLaughlin, 1992). Consequently, teachers do not need to demand respect during the lesson or spend time on classroom management but can focus on instruction as they are backed up by social norms. Furthermore, students are seen as the most important resource in Japanese societies and teaching and teachers are highly valued in accordance with that. For Chinese classrooms, Stevenson & Lee (1995) found a similar pattern,

but the teacher proceeded faster and spent more time on direct instruction and less on group or individual work compared to the Japanese teachers. A sequence of three steps (instruction-practice-feedback) was frequently being used in both countries and showed the teacher-centred approach used. As Sato & McLaughlin (1992) point out, Japanese teachers tend to support a constructivist style of teaching, which assumes that students hold responsibility for their learning. In order to maintain these high standards, Japanese teachers are required to frequently take part in professional development.

One particular feature of Japanese schools is that teachers actively take part in school policy making and describe the relationships with colleagues as very supportive and collegiate. This is, for example, represented by rotating responsibilities for school duties and the teachers share or a common room that provides working and social space (Sato & McLaughlin, 1992).

Malaysia stand out compared to the other three countries in the cluster (South Korea, Singapore and Japan), as it is a very heterogeneous country in terms of ethnicity and religions (i.e. seven different religious groups and various ethnic minorities). To develop the country, 'Vision 2020' is directing the education system, which aims at developing a 'well-rounded' human being that is economically successful, morally strong and part of a developed, secure and united Malaysian society (Lee, 1999). The particular challenge in this country is to develop national unity among different ethnic groups (i.e. Muslims and Hindus). Thus, a national curriculum and lessons in the main language of the country have been employed. Nonetheless, one of the main aims is the economic and industrial development of the country through educational excellency. Actions taken are – despite the national curriculum and the education ministry providing guidelines – decentralisation of school leadership so the principals can implement directions according to the school's needs (i.e. diversity, location) and teacher empowerment by, for example, giving them a voice in decision making processes and creating subject councils that bring forward any needs or challenges faced by the subjects taught. It might be worthwhile mentioning that – at least in the secondary school sector – private schools are rather rare, and education is mainly being delivered by state schools (ibid.).

5.4.4 SUMMARY

As can be seen from the description of the education systems in the ESEA cluster, Malaysia, in a way, stands out from the other three countries (South Korea, Singapore and Japan), that seem to be more homogeneous than Malaysia with its differences in religious beliefs and ethnic groups. Nonetheless, Rosker (2016) argues that despite the religious and ethnic plurality in Malaysia, which is different from the other three countries in the cluster, Confucianism as a 'system of ideas' that shapes the understanding of public values and an ethical system across all East- and South-East Asian countries. Accordingly, it can be assumed that despite their apparent differences, those countries are still under a 'Confucian umbrella', which helps to unite them in terms of cultural values. As Ravitch (2009) showed in her review of the social status that teachers have internationally, all Asian countries were found to highly respect teachers, which was, for instance, represented by the public discourse about teachers or a comparatively high salary.

5.4.5 TEACHING AND EDUCATION IN THE NORDIC COUNTRIES

Especially in the wake of the PISA studies, where Nordic countries (with Finland at the top) have produced above-average results and the search of a 'Nordic model of education' has started to find ways to explain the success (Bulle, 2011). Although the Nordic countries share certain similarities like a democratic approach to schooling or a strong focus on educating pupils for a civic society (Antikainen, 2006), there are more similarities between Norway, Sweden and Denmark than Finland and the other three countries, which share more similarities and closer relationships historically (Telhaug et al., 2006). Therefore, where necessary, a distinction between Finland and Norway, Sweden and Denmark will be made.

Looking at the general underlying philosophy of the education systems in the Nordic countries reveals that the underlying school philosophy is progressive and social-oriented, with a strong focus on pupil-oriented teaching strategies. This is particularly the case for Norway, Sweden and Denmark, where private schools have long been rejected, as schools are seen as a place for children to be educated as members of a civic society in an egalitarian way, regardless of their social background. Thus, private schools tend to be rejected in favour of more egalitarian

comprehensive schools. For Norwegians, for instance, 'solidarity' is a central value to be conveyed through schooling and the principals are encouraged to promote a unified culture at school that emphasises respect and tolerance. Furthermore, in Norway and Sweden a particular emphasis is put on a national curriculum that sets clear goals for schools and teachers according to Telhaug et al. (2006).

In Denmark, in contrast, an emphasis is put on neo-liberal thinking: Schools are encouraged to be independent and are given much freedom of choice in finding their own pedagogic aims and approaches towards education (Telhaug et al., 2006). The pupil in this context is seen as a 'cultural, religious and spiritual human being' (ibid. p. 273) whose personal and intellectual development should mainly flourish through knowledge and experience. Thus, there is a broader variety and choice in terms of the pedagogical values and school characteristics that contribute to the children's development than in the other Nordic countries. In Finland in particular, for a very long time a traditional and academically oriented education system has been in place that has only begun to be to be replaced in the more recent years. It seems to be particularly noteworthy that due to the academic emphasis that is also placed on teacher training and professionalisation, the teaching profession is highly regarded in Finland (Telhaug et al., 2006).

Moreover, teacher education in the Nordic countries has always been treated as an academic discipline and a high emphasis has been put on the training of teachers at universities and the professionalisation of teacher education (Westbury et al., 2005). Thus, the teaching profession and therefore teachers are highly regarded.

When it comes to classroom practices the aim to educate citizens for an egalitarian society informs those, too. For example, in Norwegian and Swedish classrooms, the teachers actively encourage students to take part in teacher-led discussions and stimulate peer learning. In Finland, in contrast, the focus is rather on teacher-centred learning and the teachers lead discussions or provoke the students' thinking by asking detailed questions and commenting or elaborating on the students' answers (Klette et al., 2018). Similarly, in Norway and Sweden a clear focus is placed on relating mathematical concepts to everyday situations, for example, whereas in Finland the understanding of concepts and a more focused classroom discourse is aimed at. Another tool to stimulate student interaction in Nordic classrooms is group work as

a complementary approach to organise the classroom and provide room for the development of democratic skills (Galton et al., 2009). Thus, the role of the teacher in the Nordic classrooms is one that not only demands the transmission of knowledge, but also feeds into wider societal aims like social justice and equality. The classroom can be seen as a micro-environment, in which these skills are developed, and a respective mindset is established.

5.4.6 SUMMARY

In the Nordic countries, the unifying characteristic seems to be a socially oriented curriculum. Despite the differences between Finland, which traditionally had a more academic approach towards education and has only started to lean towards a more progressive system in recent years and Norway or Sweden, where emphasis is placed on applying progressive strategies to instruction, the countries seem to share the main goal i.e. to educate citizens for a democratic society. Moreover, public schools enjoy a high reputation in all Scandinavian countries, as they are considered an integral part of society. With this, in these countries a high reputation of teachers comes along.

5.5 COMPARATIVE SUMMARY

The previous section had three main aims: Firstly, it sought to explain the grouping of countries in clusters with respect to features of the culture and school system, which help distinguishing those clusters from each other and thus provide a framework for interpreting the differences in teachers' self-efficacy.

Secondly, it aimed at showing the underlying values that are unique to or, at least, predominant in each country cluster and therefore illuminate some of the differences between them. These values can be seen as the philosophical underpinnings that shape the ideas of how a society should operate and the role of the education system in it.

Thirdly, it aimed at giving an overview of classroom practice and in particular the role of the teacher in the school system in different clusters and within cluster country variation.

As can be seen from the description of the underlying values that prevail in each of the country clusters, these values differ from each other to some extent but nonetheless also show

some commonalities. There are more apparent similarities between the Anglo-Saxon and the Nordic cluster, which both support rather individualistic cultures, have a rather low degree of power distance and approximately the same degree of gender egalitarianism. The ESEA countries, in contrast, support a lower degree of gender egalitarianism; which means they put more emphasis on the difference between men and women, are more collectivistic and support steep hierarchies that demand respect for the elder or superiors. The only exception found with only very little difference among the three country clusters is performance orientation, which is equally supported in all three.

The teaching styles and the role of the teachers, in contrast, behave slightly differently, as there seem to be fewer differences as between the leadership styles and country values. For example, there are fewer differences between the classroom teaching styles in the Anglo-Saxon and ESEA countries than might be expected with respect to the different aims and traditions on which they are based. In both country clusters, a mixture of teacher-centred and group style whole class teaching was applied; although to varying extents. But from the discussion of contextual factors, the aims and underlying assumptions that lead to those decisions, are rather different. While in the ESEA countries the underlying norms are based on Confucianism and thus striving for excellence, imitation of the teacher and personal development of the pupils is to be achieved, in the Anglo-Saxon countries the education systems seem to be more informed by the ideas of Dewey. Dewey suggested democratic education, the synthesis of theory and praxis and an approach to education that is inclusive across the ability scale. In this setting, the role of the teacher is less authoritarian than in the ESEA countries and in combination with the lesser degree of accepted power distance, the classroom setting in both country clusters can be expected to be very different, as the teachers' opportunities and power to control students in the Anglo-Saxon countries seem to be fewer. In the Nordic and the Anglo-Saxon countries, in contrast, the situation seems to be rather similar: In both countries, the aim is to educate future citizens with a rather democratic approach and self-directed learning is valued. Another similarity is that hierarchies are rather flat, whereas in the Nordic countries gender egalitarianism is more supported than in the Anglo-Saxon countries. What is, again, similar in all country clusters is an emphasis on academic teacher training to assure high content standards in the classroom, which is in line with the emphasis on performance orientation in each cluster.

It has nonetheless to be mentioned that these characterisations are ideal-typical descriptions of the situation in the country clusters. As can be seen from the description above, there is also within-cluster variation with respect to culture and education systems despite the commonalities they share. As such, they provide an overview and interpretive framework.

CHAPTER 6 METHODOLOGY

This chapter will provide measurement invariance tests for each sub-scale of TSE to show the validity of the construct within in each cluster. Furthermore, it contains the measurement invariance tests for constructivist beliefs about teaching, classroom climate and the instructional leadership style of the principal. Finally, it describes the method of analysis used to compare the three country clusters. In addition, Cronbach's alpha will be reported in Appendix 3 to show that the scales yield sufficient internal consistency.

6.1 MEASUREMENT INVARIANCE FOR TSE

To analyse and compare the three country clusters, three new variables were coded that contain all countries belonging to the respective clusters: An Anglo-Saxon variable with the USA, England and Australia, a Nordic variable with Denmark, Finland, Norway and Sweden and an East- and South-East Asian variable with Japan, South Korea, Singapore and Malaysia. Thus, no distinction of group membership other than the three country clusters are made in the subsequent analysis. All data was taken from the multiply imputed data set described in Section 4. In order to being able to group the respective countries, they have to at least perform metric invariance; thus, the intercepts of the items have to be equal within the suggested groups. Scherer et al. (2016) in an exploratory study could show that for a single instrument, where all 12 TSE items are used together, the Nordic countries, metric invariance could be reached, for the East and South-East Asian countries partial scalar invariance was reached and for the Anglo-Saxon countries, strict invariance was found (for a more detailed description of the different levels of invariance see Section 3). Going beyond Scherer's analysis, for the present study, further measurement invariance analysis were conducted by applying the tripartite division into teacher self-efficacy in instruction, student engagement and classroom management. The aim was to show that for each country cluster for each construct at least metric invariance can be reached, which is the minimum requirement a latent factor has to have to fulfil in order for it to be used as a latent mean in statistical analysis (see Section 4). The results can be seen in the following table:

Table 22 Measurement invariance for configural and metric invariance for ETSE, ITSE and MTSE for all three country clusters

	Anglo-Saxon countries			ESEA countries			Nordic countries		
Model fit indices	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ
SRMR_t	.047	.059	.012	.051	.050	-.001	.054	.052	-.002
SRMR_s	.046	.059	.013	.021	.060	.039	.051	.069	.018
RMSEA	.061	.060	-.001	.064	.065	.001	.056	.059	.003
CFI	.911	.921	.010	.943	.933	-.010	.913	.905	-.008
TLI	.932	.926	-.006	.949	.943	-.006	.933	.920	-.013

The data shows that for all three country clusters at least metric invariance could be reached. As the necessary level of MI was reached, no further analyses were conducted to show whether also stricter invariance constraints would hold or not. For the Anglo-Saxon countries, the SRMR_t[13] for configural MI was .047 and for metric MI .059 (Δ +.012), the SRMR_s was .047 for configural and for metric MI .059 (Δ + .012) and the RMSEA was .061 for configural and .060 for metric MI (Δ - .001). The CFI rose from .911 for configural to .921 for metric MI (Δ + .010) and the TLI dropped from .932 for configural to .926 for metric MI (Δ - .006). For the East and South-East Asian countries, the SRMR_t dropped from .051 (configural MI) to .050 (metric MI) (Δ - .001), the SRMR_s rose from .021 for configural MI to .060 for metric MI (Δ + .039) and the RMSEA rose from .064 for configural to .065 for metric MI (Δ + .001). For both, the CFI and the TLI, the value dropped after adding constraints to the model: The CFI dropped from .943 for configural to .933 for metric MI (Δ -.010) and the TLI dropped from .949 for configural to .946 for metric MI (Δ -.010). Lastly, for the Nordic countries the SRMR_t dropped from .054 (configural MI) to .052 (metric MI) (Δ -.002) and the SRMR_s rose from .051 (configural MI) to .069 (metric MI) (Δ + .018). The RMSEA, in contrast, rose from .056 (configural MI) to .059 (metric MI) (Δ +.003). Both, the CFI and TLI dropped after adding further constraints: the CFI dropped from .913 for configural to .905 for metric MI (Δ -.008) and the TLI from .933 for configural to .920 for metric MI (Δ -.013).

Furthermore, only a few of the changes in the goodness of fit-indicators exceed the cut-off value of Δ +/-.03 for a significant change in the goodness of fit after adding invariance constraints (see Section 3.6.2 for further explanations) and also none of the absolute goodness

[13] _t denotes the SRMR on the teacher level, _s the SRMR on the school level

of fit values were exceeded. Thus, using the tripartite structure of TSE in each of the country clusters for further statistical analysis is therefore deemed acceptable.

6.2 MEASUREMENT INVARIANCE FOR CONSTRUCTIVIST BELIEFS ABOUT TEACHING

The following section shows the results of the measurement invariance analysis for constructivist beliefs about teaching. As previously, the highest level of invariance needed in the structural equation modelling framework is metric invariance, which is why no higher level was tested.

Table 23 MI for configural and metric invariance for constructivist beliefs about teaching

Model fit indices	Anglo-Saxon			Nordic			ESEA		
	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ
SRMR_t	.011	.046	.035	.021	.052	.031	.025	.107	.082
SRMR_s	.068	.077	.009	.058	*.084*	.026	*.082*	*.095*	.013
RMSEA	.025	.106	.081	.044	.080	.036	*.088*	.046	-.042
CFI	.997	.990	-.007	.983	.899	-.084	.970	.907	-.063
TLI	.991	*.892*	-.099	.948	*.826*	-.122	.910	*.841*	-.069

As the table above shows, metric invariance could be reached in all three country clusters; only a few values slightly exceeded the respective cut-off values[14]. In the Anglo-Saxon countries, the initial SRMR for configural invariance was .011 and rose to .046 when metric invariance was tested (Δ +.035). The RMSEA was .025 for configural MI and rose to .106 for metric MI (Δ +.081) and thus exceeded the cut-off slightly. The CFI changed from .997 (configural MI) to .990 (metric MI) (Δ -.007) and the TLI from .991 (configural MI) to .892 (metric MI) (Δ -.099). Here again, the cut-off has slightly not been met.

In the Nordic countries, the SRMR was .021 for configural MI and rose to .052 for metric MI (Δ +.031). The RMSEA started at .044 for configural and rose to .080 for metric MI (Δ +.036). The CFI was .983 for configural MI and dropped to .899 for metric MI (Δ -.084), similar to the TLI, which was .948 for metric MI and came to .826 for configural MI (Δ -.122).

[14] CFI/TLI >.900, SRMR/RMSEA < .080

Finally, in the ESEA countries, the SRMR started at .025 (configural MI) and rose to .107 (metric MI) (Δ +.82) after adding constraints. The RMSEA was .088 for configural MI and then dropped to .046 for metric MI, after constraints were added (Δ -.042). The CFI was initially .970 (configural MI) and then dropped to .907 (metric MI) (Δ -.063) after constraints were added, whereas the TLI started at .910 for configural MI and then dropped to .841 for metric MI (Δ -.069). As can be seen, not all values met the strict cut-off values of .900 for the CLI and TLI or .080 for the SRMR and RMSEA. However, even after adding constraints to the model, only one outlier tended to occur. Thus, the scale is deemed acceptable for the use of comparing the three country clusters.

6.3 MEASUREMENT INVARIANCE FOR CLASSROOM CLIMATE

The following section will show that at least metric invariance was reached in all three country clusters regarding the classroom climate and thus the data fulfil the criteria for being used in the SEM context.

Table 24 2 MI for configural and metric invariance for disciplinary classroom climate

Model fit indices	Anglo-Saxon			Nordic			ESEA		
	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ
SRMR_t	.013	.021	.008	.004	.022	.018	.004	.016	.012
SRMR_s	.011	.020	.009	.059	.100	.041	.021	.060	.039
RMSEA	.039	*.101*	.062	.034	.116	.082	.031	*.072*	.041
CFI	.997	.966	-.031	.997	.948	-.049	.998	.981	-.017
TLI	.991	.941	-.050	.992	.910	-.082	.994	.968	-.026

In the Anglo-Saxon countries, the SRMR for configural MI was .013 and then rose to .021 when metric MI constraints were added (Δ +.008). The RMSEA started at a slightly higher value, .039, for configural MI and then rose to .101 for metric MI (Δ +.062). The CFI was .997 (configural MI) and dropped to .941 (metric MI) after adding invariance constraints (Δ -.050).

For the Nordic countries, the initial SRMR was .004 (configural MI) and rose to .022 (metric MI) (Δ +.018). The RMSEA also rose between configural and metric MI from .034 to .116 ((Δ +.082). Here, the intended cut-off of .080 was exceeded. The CFI started relatively high at .997 and then dropped to acceptable .948 between configural and metric invariance (Δ -.049). Similarly, the TLI started at .992 (configural MI) and then came to acceptable .910 (metric MI) (Δ -.082).

Finally, also in the ESEA countries, at least metric MI could be achieved. The SRMR started at .004 and then rose to .016 from configural to metric MI (Δ +.012). The RMSEA started at acceptable .031 (configural MI) and came to .072 (metric MI) after adding constraints (Δ +.041). The CFI started at .998 in the configural model and came to .981 in the metric model (Δ -.017). In a similar way, the TLI dropped from .994 (configural model) to .968 (metric model). Here, both times the cut-off of .900 was met.

Similar to the invariance tests for constructivist beliefs about teaching, almost all parameters stayed below the .080 cut-off value for the RMESA and SRMR and above the .900 cut-off value for the CFR and TLI when applying constraints to test if metric invariance holds. Thus, it can be assumed it is appropriate to use the scale measuring a good classroom climate for the intended analysis to compare predictors of three types of teacher's self-efficacy across three country clusters.

6.4 MEASUREMENT INVARIANCE FOR INSTRUCTIONAL LEADERSHIP

This section contains the result for the measurement invariance testing for the instructional style of leadership of the principals and shows the suitability of the scale for the proposed research.

Table 25 MI for configural and metric invariance for instructional leadership style

Model fit indices	Anglo-Saxon			Nordic			ESEA		
	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ	Config. IV	Metric IV	Δ
SRMR	.000	.021	.021	.000	.022	.022	.000	.056	.056
RMSEA	.000	*.068*	.068	.000	.037	.037	.000	.024	.024
CFI	1.000	.991	-.009	1.000	.938	-.062	1.000	.978	-.022
TLI	1.000	.981	-.019	1.000	*.876*	-.124	1.000	.975	-.025

Most notably, the unrestricted model testing configural invariance yielded perfect model fit within all three country clusters, showing values of .000 for the SRMR and RMSEA as well as values of 1.000 for the CFI and TLI. When adding invariance constraints, for the Anglo-Saxon countries, the SRMR rose to .021 and the RMSEA to .068. The CFI dropped to .991 and the TLI to .981 for metric invariance.

In the Nordic countries, when testing if metric invariance holds the SRMR rose to .022 and the RMSEA to .037. Likewise, the CFI dropped to .938 and the TLI to .876.

Finally, adding constraints to test for metric invariance in the East- and South-East Asian countries led to a SRMR of .056 and a RMSEA of .024. The CFI came to .978 and the TLI to .975.

This analysis shows that apart from two values (metric RMSEA in the Anglo-Saxon countries and metric TLI in the Nordic countries) that the scales meets the criteria for the level of metric invariance. Hence, using the scale for the following analysis can be deemed appropriate.

6.5 ANALYTICAL STRATEGY

The research questions were then analysed using the GROUPING-command in MPlus. Hereby, known groups (i.e. the three country clusters) are specified to perform a comparative analysis between the groups (c.f. MPlus User's guide V8). For each of the three domains of teacher's self-efficacy, a separate model was run. Statistically, it is assumed that all factor loadings of the latent variables within the clusters are equal, which has just been shown beforehand.

The model build-up is similar to the one of research question one (see Chapter 4), with a slight difference. Instead of adding the sets of variables separately, after calculating the baseline model all predictors were added at once. This decision was made as the main interest was to compare predictors across the three country clusters, instead of investigating the different amount of variation in teacher's self-efficacy that could be explained by adding another set of variables. Formally stated, research question 2 (country cluster analyses) is as follows:

Equation 13 Full model RQ2: Teacher-, Classroom-, Principal- and School predictors

$$TSE_{ij} = \beta_{0ij} + \beta_{BLOCK1} Teacher\ characteristics_{ij} + \beta_{BLOCK2} Classroom\ characteristics_{ij} + \beta_{BLOCK3} Principal\ characteristics_{j} + \beta_{BLOCK4} School\ characteristics_{j} + u_{0j} + e_{ij}$$

with:

i = teacher, j = school

β = predictor

u_{0j} = random term on the school level

e_{ij} = random term on the teacher level

CHAPTER 7 RESULTS RESEARCH QUESTION 2: INTERCULTURAL DIFFERENCES BETWEEN PREDICTORS OF TEACHER'S SELF-EFFICACY

This chapter presents the results for research question (RQ) 2. This research question expands on the aspects studied in RQ1 by focusing more specifically on variation for three country clusters. Here, the four sets of predictors (teacher, classroom, principal and school related) tested originally across 32 OECD and partner countries were again employed in multilevel models now comparing and explaining variation among three distinct country clusters. The first cluster, the Anglo-Saxon Cluster, contains three countries: the USA, England and Australia. The second cluster, with Nordic countries, consists of Denmark, Finland, Norway and Sweden. Finally, the third cluster of East- and South East Asian countries contains four countries - Japan, South Korea, Singapore and Malaysia. The three country clusters were identified from the same data set as that used to explore the first research question. The multilevel models accordingly, were run using the same strategy with 5 multiple imputed data sets in MPlus. The information regarding the structural validity of the latent variables (ETSE, ITSE, ETSE, constructivist beliefs about teaching and classroom climate) can be found in Sections 6.1 – 6.4.

The clusters consist of the following number of teachers as outlined in Figure 1-1:

Table 26 Number of teachers and schools in each country cluster

Country	n teachers	n schools
Anglo-Saxon Country Cluster	6,477	297
Nordic Country Cluster	10,678	620
East & South-East Asian Country Cluster	12,510	678

Models were run separately for each of the three country clusters and are reported in separate sections. The section deals with one country cluster and is divided into three sub-sections according to the three aspects of teachers' self-efficacy (ITSE, ETSE and MTSE) that are the focus of this research. At the end, the differences across the countries for each dimension of TSE will be compared.

7.1 RESEARCH QUESTION AND HYPOTHESES

As with the first research question, which compared predictors of teachers' self-efficacy across 32 countries, the same predictors will now be analysed within the three country clusters mentioned above. There is one main research question guiding the analysis:

To what extent can teacher self-efficacy be explained by teacher-, classroom-, principal- and school characteristics within each country cluster?

Based on previous research (discussed previously in the literature review in Chapter 2), the following hypotheses were also stated and investigated:

1. Male teachers have a higher sense of TSE
2. More experienced teachers have a higher sense of TSE
3. Teachers who teach higher performing students have a higher sense of TSE
4. A warmer and more organised classroom climate is related to a higher sense of TSE
5. Role-model behaviour of the principal (e.g. in instructional leadership) is related to a higher sense of TSE

7.1.1 TO WHAT EXTENT IS TSE IN INSTRUCTION PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN EACH OF THE THREE COUNTRY CLUSTERS?

This section presents the results for models predicting teachers' self-efficacy in instruction (ITSE) for each of the country clusters, beginning with the Anglo-Saxon cluster, followed by the Nordic cluster followed by the East- and South-East Asian cluster. The section will be concluded with an overview of the differential results and a comparative summary. The results section for each cluster is merely descriptive and all interpretation and discussion of the results is left to the comparative section.

At first, an empty model with no predictors included was calculated against which the full model can then be compared. As the model fit information is the same for all three country clusters, it will only be reported once in the introduction to the section, but not repeated in the results section of each cluster. The initial 2-log likelihood was 214,489.18. By adding the predictors, the model fit could be improved as the -2log likelihood was reduced to 211,360.43 (Δ -3,129).

7.1.1.1 Anglo-Saxon countries

In the Anglo-Saxon countries, among schools, the initially unexplained variance was 3.0% and among teachers 97.0%. The unexplained variance among teachers was reduced by 2.2% after the addition of predictors, but no more variance could be explained among schools by adding teacher, classroom, principal and school characteristics.

Considering results for the predictors pertaining to teacher characteristics, all except the educational level were positively associated with ITSE. Female gender showed a positive relationship with ITSE (β = .077, p < .001, ES = .67) as did teaching experience (β = .003, p < .001, ES = .03). The amount of time per week spent on teaching (β = .001, p < .01, ES = .01), spent on planning lessons (β = .003, p < .01, ES = .03) and spent on administrative duties (β = .004, p < .001, ES = .03) only accounted for a very small proportion of the variation in ITSE. Constructivist beliefs about teaching also showed a weak association with ITSE (β = .030, p < .001, ES = .26).

From the eight classroom characteristics, four showed a significant association with ITSE: A good classroom climate was strongly positively associated (β = .027, p < .001, ES = .23), as well as low student achievement (β = .012, p < .01, ES = .10). Only the time allocated to administrative tasks (β = -.004, p < .01, ES = -.03) was associated with ITSE, but not the time spent on teaching and learning or keeping order. Finally, the percentage of students from a disadvantaged social background (β = .026, p < .001, ES = .23) showed a positive association with ITSE. This can be similarly explained as the positive association between ITSE and low student achievement: the teacher's mastery experience is the higher, the more there is room for improvement on the students' behalf. It is nonetheless somewhat surprising, as this also tends to come along with more challenging classroom situations.

None of the principal characteristics show a positive association with ITSE, but all three school characteristics: School enrolment numbers are very weakly positively related to ITSE (β = .000, p < 05., ES = 00[15].), status as a private school negatively related (β = -.012, p < .01, ES = -.10) and the community size again positively (β = .012, p < .01, ES = 10).

[15] These numbers are smaller than decimal points displayed by MPlus, which is why they appear 0. Thus, they are considered tiny.

Table 27 Results for TSE in instruction for all three country clusters

	Anglo-Saxon Cuntries							Nordic Countries							East- & South-East Asian Countries						
	Baseline Model			Full model				Baseline Model			Full model				Baseline Model			Full model			
Response	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**
Fixed Part																					
intercept	3.285	***	.016	3.285	***	.016		3.285	***	.016	3.285	***	.016		3.285	***	.016	3.285	***	.016	
Teacher																					
Female Gender				.077	***	.010	.67				.075	***	.008	.69				.035	***	.008	.25
Educational Level				.025	n.s.	.025	.22				.005	n.s.	.012	.05				.026	*	.014	.19
Teaching Exp.				.003	***	.001	.03				.002	***	.000	.02				.005	***	.000	.04
Hours Teaching				.001	**	.001	.01				.004	***	.001	.04				.002	***	.001	.01
Hours Planning				.003	**	.001	.03				.003	**	.001	.03				.001	n.s.	.001	.01
Hours Admin. Work				.004	***	.001	.03				.005	***	.001	.05				.001	**	.001	.01
Construct. Bel.				.030	***	.003	.26				.037	***	.002	.34				.048	***	.002	.34
Classroom																					
Class size				-.001	n.s.	.001	-.01				.002	***	.001	.02				.000	n.s.	.000	.00
Classroom Climate				.027	***	.003	.23				.026	***	.002	.24				.026	***	.002	.19
Low Student Achievement				.012	**	.006	.10				.023	***	.005	.21				-.008	n.s.	.005	-.06
Teaching & Learning				.000	n.s.	.001	.00				-.001	n.s.	.001	-.01				.003	**	.001	.02
Keeping Order				.000	n.s.	.001	.00				-.001	n.s.	.001	-.01				.002	**	.001	.01
Admin. Tasks				-.004	**	.002	-.03				-.002	n.s.	.001	-.02				.004	***	.001	.03
Different Language				.004	n.s.	.005	.03				.024	***	.005	.22				.026	***	.004	.19
Disadv. Background				.026	***	.006	.23				.022	***	.006	.20				.051	***	.005	.36
Principal																					
Female Gender				.006	n.s.	.013	.05				.001	n.s.	.014	.01				.080	***	.016	.57
Age				-.001	n.s.	.001	-.01				.001	n.s.	.001	.01				-.010	***	.001	-.07
Educational Level				-.025	n.s.	.020	-.22				-.051	*	.027	-.47				-.001	n.s.	.040	-.01
Work Experience				.000	n.s.	.020	.00				.000	n.s.	.001	.00				.005	**	.002	.04
Instructional Leadership				.002	n.s.	.003	.02				.006	*	.004	.06				.046	***	.004	.33
School																					
Student Enrolment				.000	*	.000	.00				.000	***	.000	.00				.000	***	.000	.00
Private School				-.012	**	.013	-.10				-.101	***	.020	-.93				-.040	n.s.	.028	-.28
Commuity Size				.012	**	.005	.10				-.007	n.s.	.006	-.06				-.039	***	.006	-.28
Random Part																					
School	.008	***	.002	.001	n.s.	.001		.023	***	.003	.008	***	.001		.042	***	.002	.138	***	.003	
Teacher	.260	***	.004	.117	***	.004		.271	***	.005	.114	***	.003		.193	***	.003	.020	***	.031	
-2*loglikelihood:	214489.18			211360.43				214489.18			211360.43				214489.18			211360.43			
Δ -2*loglikelihood				-3,129							-3,129							-3,129			
% variance				Explained variance				% variance			Explained variance				% variance			Explained variance			
School	3.0%			School	n.s.			School	7.8%		School	n.s.			School	17.9%		School	79.5%		
Teacher	97.0%			Teacher	2.2%			Teacher	92.2%		Teacher	1.4%			Teacher	82.1%		Teacher	n.s.		
ICC																					
School	3.0%	**	.008	0.8%	***	.008		7.8%	***	.011	6.6%	***	.010		17.9%	***	.014	87.3%	***	.011	
Teacher	97.0%	***	.008	99.2%	***	.008		92.2%	***	.011	93.4%	***	.010		82.1%	***	.014	12.7%	***	.011	

7.1.1.2 Nordic countries

The second set of results concerns the Nordic countries. Among schools, the initially unexplained variance proportion was 7.8% and among teachers 92.2% in the Nordic countries. By adding the predictors among teachers 1.4% additional variance could now be explained.

All but one (educational level) of the suggested teacher characteristics contributed significantly to explaining variation in ITSE. Female gender showed a positive association with TSE (β = .075, p < .001, ES = .69) as did teaching experience (β = .002, p < .001, ES = .02). All three elements of time allocation during the week were also positively, though very weakly, related to ITSE: Hours spent on teaching (β = .004, p < .001, ES = .04), hours spent on planning lessons (β = .003, p < .01, ES = .03) and hours spent on administrative work (β = .005, p < .001, ES = .05). Finally, constructivists beliefs about teaching a teacher has also shows a moderate positive association with ITSE (β = .037, p < .001, ES = .34).

From the classroom characteristics, all but the percentage of classroom time spent on teaching and learning, keeping order or administrative tasks were significant but weak predictors of variation in ITSE scores in Nordic countries.

A bigger class size showed a very weak but positive association (β = .002, p < .001, ES = .02), as well as a good classroom climate (β = .026, p < .001, ES = .24). Student related features like low achievement (β = .023, p < .001, ES = .21), a different mother tongue than the language of instruction (β = .024, p < .001, ES = .22) and a disadvantaged social background (β = .022, p < .001, ES = .20) were also positively related to ITSE.

Out of the five suggested principal characteristics, two showed a weak but significant relationship with ITSE: Educational level negatively, but with a moderate ES (β = -.051, p < .05, ES = -.47) and an instructional leadership style positively (β = .006, p < .05, ES = .06).

From the three characteristics describing the school, two significantly predicted ITSE in Nordic teachers: The number of students enrolled (β = .000, p < .001, ES = .00) and teaching at a private school (β = -.101, p < .001, ES = -.93).

7.1.1.3 East- and South-East Asian countries

Thirdly, the results for the East- and South-East Asian country cluster are presented. In the East- and South-East Asian countries, 17.9% of the initially unexplained variance resided among schools and 82.1% among teachers. After adding the predictors in the full model, among schools further 79.5% of the variance could be explained, but not more among teachers.

From the teacher characteristics, all but one (hours per week spent on planning lessons) showed a positive association with ITSE. Against my hypothesis, female gender was moderately strongly related to a high sense of ITSE (β = .035, p < .001, ES = .25), as well as the teacher's educational level β = .026, p < .05, ES = .19) and teachers' constructivist beliefs about teaching (β = .048, p < .001, ES = .34). Teaching experience (β = .005, p < .001, ES = .04), hours per week spent on teaching (β = .002, p < .001, ES = .01) and hours spent on administrative work (β = .001, p < .01, ES = .01) were only very weakly associated with ITSE.

A good classroom climate showed a weak positive association with ITSE (β = .026, p < .001, ES = .19), as the percentage of students with a different mother tongue than the language of instruction (β = .026, p < .001, ES = .19) and the percentage of students from a disadvantaged background, but here the ES was moderate (β = .015, p < .001, ES = .36).

The percentage of classroom time spent on teaching and learning (β = .003, p < .01, ES = .02), on keeping order (β = .002, p < .01, ES = .01) or on administrative tasks (β = .004, p < .001, ES = .03) only had a very small association with ITSE. In contrast to my hypothesis, student achievement was not significantly related to ITSE in the ESEA countries.

Again, in contrast to the other two country clusters, most of the principal characteristics were significantly related to the teachers' sense of efficacy in instruction. Female gender (β = .080, p < .001, ES = .57) and instructional leadership (β = .046, p < .001, ES = .33) showed a moderately strong association with ITSE, whereas the principal's age (β = -.010, p < .001, ES = -.07) and work experience (β = .005, p < .01, ES = .04) were only weakly related to a higher sense of ITSE.

From the school characteristics, a higher number of students (β = .000, p < .001, ES = .000) showed a negligible association with ITSE and the size of the community in which

the school is located a stronger but still relatively small negative association (β = -.039, p < .001, ES = -.28).

7.1.2 Similarities and differences in the country clusters for teachers' self-efficacy in Instruction

This section will highlight the differential results for predictors of teachers' self-efficacy in instruction across the three country clusters. It is divided into four parts, describing the differences for teacher-, classroom-, principal- and school characteristics separately. The main means by which those differences will be explored relate to possible country differences in terms of cultural values supported in the countries and specific characteristics of the education systems, but with respect to previous research on TSE. The effect sizes will be reported as an unbiased measure of the magnitude of the differences, as they are independent of the sample size of the respective clusters.

7.1.2.1 Teacher characteristics

At first, the differences in teacher characteristics will be reported. Part of the analysis were teacher gender, educational level, work experience, hours per week spent on teaching, planning lessons or administrative duties and the constructivist beliefs about teaching a teacher holds.

For female teacher gender, the association in the Anglo Saxon (ES = .67) and Nordic (ES = .69) cluster were rather similar, but much smaller in the East- and South East Asian cluster (ES = .25). This means that being a female teacher in the Anglo-Saxon or Nordic countries predicted a higher level of ITSE than being a female teacher in an East- or South Asian country. As found by House et al. (2004), the support for gender egalitarianism is slightly higher in the Anglo-Saxon and Nordic countries than in the ESEA countries. In gender egalitarian countries the support for ambitious women and female teachers, as such, tends to be higher, which could explain the difference here. Previous research on female gender and TSE, however, has produced mixed results so far, where Cousins et al. (1996) found a positive association for female gender and TSE, whereas other studies have found no gender association (i.e. Malmberg et al., 2014) or a higher TSE for male teachers (Coladarci et al., 1992). Thus, the differential results here are supported by previous research in both ways. It might be noteworthy that those

results were found in Anglo-Saxon classrooms throughout. Thus, in the present study the explanations for the differential results of the country clusters might, rather, be found in the cultural values as for example gender egalitarianism here, which prevails in these countries, than in the contexts of previous studies on TSE.

The teachers' educational level and its association with ITSE was only significant in the ESEA cluster. This accords with previous research (Fackler & Malmberg, 2016; Cousins et al., 2016). Especially for teachers in a very high-demanding environment driven by Confucian values, where teachers are considered experts (Pang, 2006; Gardner, 2014), a higher level of education could be particularly important, as it adds to the teacher's seniority and credibility with respect to content knowledge.

For the years of work experience, the Anglo-Saxon and Nordic cluster showed a similarly strong relationship with ITSE, whereas the relationship in the ESEA countries was slightly stronger, but the effect sizes were rather small, ranging from ES = .02 – ES = .04 and thus explanations will only be made tentatively. Previous research, however, found mixed results: For instance, Chester & Beaudin (1996) & Fackler & Malmberg (2016) found that years of work experience were associated with a higher level of TSE, whereas Klassen & Chiu (2010) found that TSE rises until mid-career and then drops and Pas et al. (2012) found no association between TSE and work experience. Years of work experience can be seen as a source of mastery experience, during which teachers can accumulate successful action strategies, but can also lead to frustration if too many setbacks have been experienced. A possible explanation for the positive relationship between ITSE and work experience in the ESEA cluster in particular could be that teachers are typically found to be more highly supported and respected throughout society (Sato & McLaughlin, 1992), which might foster feelings of ITSE.

In the Nordic countries (ES = .04), the relationship for the number of weekly hours spent on teaching was a little stronger than in the Anglo-Saxon and in the East- and South-East Asian countries (both times ES = .01). Out of the duties the teacher has to perform in the classroom, teaching, as such, is the one most closely related to TSE in instruction. Hence, it seems surprising that across all countries, the effect size is very small.

Between Anglo-Saxon and Nordic countries (ES = .03), there was no differences in the magnitude of hours spent on planning lessons and in the ESEA countries, the association was not significant. Thus, putting more hours of lesson planning predicted a higher ITSE in Anglo-Saxon and Nordic countries than in East/South Asian ones. As for example McClelland, Rindlisbacher & DeCharms (1995) have shown, protestant work ethics as are found in Anglo-Saxon or Nordic countries, lays emphasis on hard work and achievement, which could possibly be reflected in the effort put into planning lessons. Given the small effect size, this explanation is given tentatively.

For hours per week spent on administrative work, the picture is similar: There is no big difference between the Anglo-Saxon (ES = .03), Nordic countries (ES = .05) and ESEA countries (ES = .01). However, putting more hours of admin predicted a slightly higher ITSE in Anglo-Saxon and Nordic, than in East/South Asian countries. Given that the ES for all three types of work allocation are rather small, as well as the differences found and thus it is difficult to make thorough interpretations; especially without a strong base of previous research.

Investigating the difference between the constructivist beliefs about teaching revealed differences between the ESEA and Nordic and the Anglo-Saxon cluster: The highest association was found in the ESEA (ES = .34) and the Nordic (ES = .34) cluster; the comparatively weakest association was found in the Anglo-Saxon cluster (ES = .26). Features of the constructivist beliefs about teaching are, for example, the view of the learner of being actively involved in the lessons and responsible for their own learning as well as that cognitive processes are important for student inquiry. This is in line with previous research, which has found that high TSE teachers employ more diverse teaching strategies (Emmer & Hickman, 1991) and accords with the preference for student-centred teaching approaches in the Nordic countries (Westbury et al., 2006). For the ESEA countries, a possible strand of explanation might be that in more teacher-centred settings the idea of students being responsible for their own learning removes some of the pressure from the teachers, as they believe they share the responsibility with their students. This has been previously found by Sato and McLaughlin (1992), which is surprising, considering that teacher input and the curriculum are key characteristics of the school systems in the ESEA countries (i.e. Gardner, 2014; Lim, 2007). However, it might explain the comparatively strong association with ITSE and constructivist beliefs about teaching

in the ESEA countries. Furthermore, in countries with a strong focus on performance orientation and a high social status of schools, these features might be regarded as important characteristics of good schooling. But it should, nonetheless, be pointed out that constructivist beliefs about teaching contributed significantly to explaining variation in ITSE in all three clusters and only the magnitude of the association with ITSE varies slightly.

7.1.2.2 Classroom characteristics

This section reports and analyses the differences found in the responses for classroom characteristics. Subjects of the analysis were class size, classroom climate, % of low achievers in the class, % of classroom time spent on administrative duties, keeping order and teaching and learning, the % of children with a mother tongue other than the language of instruction and % of children from disadvantaged homes.

Only class size showed a significant association in the Nordic countries (ES = .02), but the effect size is negligible.

None of the associations of the classroom climate with ITSE in the three country clusters differed largely from each other (ES = .19 - ES = .24). Thus, it can be assumed that a good classroom climate contributes approximately equally to explaining ITSE in all three country clusters regardless of cultural differences and can be seen as an important feature of successful teaching in relation to predicting a more favourable TSE. It can be assumed, however, that teachers in classrooms where a good atmosphere is maintained, are likely to experience success in their teaching and thus have a source of mastery experience. Furthermore, this measure of classroom climate suggests also that the students themselves take care of a good classroom environment and thus teachers also have a source of verbal persuasion, as they see their efforts paying off. It also provides an environment that allows the teacher to focus on instruction.

The percentage of low achievers has a positive association in the Anglo-Saxon (ES = .10) and Nordic (ES = .21) countries but was not significant in the ESEA cluster. This suggests that having a larger proportion of lower achieving students in a class seems to be associated with a small boost for teachers' ITSE in the Nordic or in Anglo-Saxon countries, but less so in Asian/South-East Asian countries. One possible source of explanation could be the different

levels of performance orientation in the country clusters. In the Anglo-Saxon and Nordic countries, performance orientation is almost equally strongly supported, but is somewhat lower in the ESEA countries (House et al., 2004). Accordingly, the ITSE of teachers in the ESEA countries might be predicted to a different degree by student achievement than in the other two country clusters. Alternatively, schooling in China as part of the ESEA countries is highly supported also through private efforts of the family, which might yield better results in general, and bad performance could lead to direct student support for better results in the future. Thus, bad student performance might be less of an issue there than in the other two clusters. For the teachers in the Anglo-Saxon and Nordic countries, the result is against my hypothesis, derived from Malmberg et al. (2014) and Ashton (1983), who found a positive relationship between student achievement and TSE, as teachers could perceive bad achievement as a failure of their own instructional abilities. However, at second sight it seems to be plausible that for low achievers there is more room for improvement, and it is easier for the teacher to allocate student success to their own effort than, for example, parental involvement, which is more likely to be the case for high achieving students in any case.

The percentage of time spent in the classroom spent on administrative work was found not to be predicting variation in ITSE in the Nordic countries but showed a positive association with ITSE in the ESEA countries (ES = .03) and a negative association in the Anglo-Saxon countries (ES = -.03). In these clusters the ES suggests only a negligible association.

Lastly, as with the time spent on teaching and learning was only found to be significant in the ESEA countries (ES = .02), given the stronger orientation towards school performance and the high demands society and parents have on teachers, this result is perhaps not surprising. Although the other countries deem schooling very important, too, the emphasis in the ESEA countries seems to exceed the relative importance in the other countries (Pang, 2011). However, this ES is again very small and thus interpretations should be made cautiously.

The following characteristics describe the social context of the classroom. The association of ITSE and the percentage of students with a different mother tongue from the language of instruction showed a significant association with ITSE in the Nordic (ES = .22) and ESEA (ES = .19) countries, but not the Anglo-Saxon cluster. This result stands out from the other results, as it suggests a greater similarity between the Nordic and ESEA cluster but not between the

Nordic and Anglo-Saxon cluster, as it has mostly been the case previously. This contradicts previous research, which found for Dutch teachers a lower level of TSE in instruction in the case of a higher percentage of minority students in the classroom (Geerlings et al., 2018). However, the language of instruction is the main means by which teachers transmit knowledge and thus a source of mastery experience. Given the history of Anglo-Saxon countries like the USA and England, which have been facing a lot of migration throughout history, teaching in multi-cultural settings has long been part of the teachers' work environment. Thus, it can be suspected that in countries where the teaching environment is traditionally less diverse, there is a stronger association between ITSE and the percentage of foreign students. It has been found by Caijkler & Hall (2009) that teachers with many students with a mother tongue different from the language of instruction have to cope with more challenges in teaching their students, but also that this can be compensated through supervision. In particular, language as a means of instruction can be learned and it could be suggested that bilingual students can achieve good grades, but this needs to be investigated in further research.

The percentage of students from a disadvantaged background was significantly related to ITSE in all three clusters but shows the largest effect size in the ESEA cluster (ES = .36 compared to ES = .23 and ES = .20). Accordingly, the relationship between ITSE and the percentage the western country clusters is weaker. Strand (2012) found that local low SES students ('white British' in his study) were outperformed by low SES students with a migrant background. According to Strand (2012) the different parental academic aspirations explained the difference, which was higher for those parents with a migrant background. Furthermore, in the Anglo-Saxon countries there is special attention paid to not let disadvantaged students fall behind (Norris, 2004). This might explain why teachers with a larger number of low SES students and students from migrant backgrounds display more confidence in their teaching abilities. However, when it comes to student SES and its association with ITSE, the ESEA countries stand out again. As already mentioned, one particular feature of the ESEA countries is their strong focus on school performance and their preference for teacher-centred lessons with only careful use of more progressive methods like group work. Also, in such a setting, students from a less affluent background are less likely to 'slip through the net' (Lim, 2007) and the teacher can recognise and then react towards individual needs more quickly than in an educational setting

where students work more individually, which tends to be the case in more progressive school systems as in the Nordic or Anglo-Saxon countries.

7.1.2.3 Principal characteristics

In the following section, the differences between the country clusters for the principal characteristics will be reported. The characteristics focused on are principal's age, work experience, educational level and instructional leadership style practised at school.

For the principal's female gender, it was neither significant in the Anglo-Saxon nor in the Nordic country clusters but showed a moderately strong positive relationship with ITSE in the ESEA cluster (ES = .57). It has been suggested that female principals tend to employ a more empathic leadership style (Eagly et al., 1992), which could be supportive of developing a higher level of ITSE in an environment that is characterised by high demands on the teachers (Pang, 2011). The principal age, in contrast, has a very weak negative association (ES = -.07) in the ESEA countries, but no significant relationship in the other two clusters. One particular feature that prevails in the ESEA countries is supposed to be greater respect for the elder, which is on the one hand part of Confucianism (Gardner, 2014). On the other hand, it is supported in societies which accept large power distance between inferiors and superiors (Hofstede, 2011), which the ESEA countries do. In such an environment, the age and the perceived superiority, which comes along with it, could also lead to greater distance that is perceived negatively. In the Nordic and Anglo-Saxon countries, which, in contrast, support rather flat hierarchies, where the relationship between superiors and other workers is rather more egalitarian. However, the effects found in the in ESEA cluster are still very weak so do not lend much support for the 'respect for the elder' cultural influence as important for TSE.

A similar pattern arises for the principal's work experience, where there is only a very weak positive association found in the ESEA cluster (ES = .04), which is almost negligible as there is no sound research base available that could help explain the result.

The educational level, in contrast, only showed a significant but negative association in the Nordic countries (ES = -.47). The principal's educational level can be seen – similarly to work experience – as a possible indicator for the principal's seniority. This result is somewhat surprising, as this is a highly respected feature in Confucian countries with a high level of power

distance, but not so much the Nordic countries. In the more egalitarian countries, in contrast, it could be that a more experienced principal who is rather demanding is perceived as intimidating or perhaps out of touch. However, this would need further research.

For the association between the principal's instructional leadership style and ITSE, there was no significant relationship in the Anglo-Saxon cluster, but a weak positive association in the Nordic (ES = .06) and a stronger effect for ESEA (ES = .33) clusters. For these two clusters, the result is in line with my hypothesis (Kirk, 2016). First of all, this leadership style comprises features that are supported in both country clusters, regardless of their cultural values. Those features are, for example, the setting of academic goals or role model behaviour of the principal, which could lead to a more supportive work environment and could give the teachers guidance and security. It could also serve as a source of vicarious experience, especially as the setting of academic goals is in line with the focus on school performance in the ESEA cluster. The role model behaviour displayed by the principal can be regarded as a source of vicarious experience, which can boost the instructional SE of teachers as it gives them a sense of what is expected of them. With respect to theory, this result can be explained in two ways: the Anglo-Saxon countries see themselves as performance oriented, which fits well with the significant association between ITSE and instructional leadership style. Conversely, Nordic countries only support a low level of power distance and teachers might expect to participate in school decisions than being told top-down. This is supported by Ashton et al. (1993), who found the opportunity for teachers to influence decisions positively was associated with TSE. As the association between leadership and ITSE was rather weak, it will only be tentatively said that there might also be pressure associated with a high level of performance orientation, which is why a directive and role-model principal might actually, nonetheless, provide a sense of security.

7.1.2.4 School characteristics

This sub-section discusses the difference in the association in ITSE and the school's location, the status as a private school and the number of students enrolled at school between the three country clusters of Anglo-Saxon, Nordic and ESEA countries.

For the size of the community in which the school is located, the strongest, but negative association was found in the ESEA cluster (ES = -.28) and a weak but positive association in the Anglo-Saxon cluster (ES = .10). The school's location can be seen as a possible indicator of school composition. In bigger cities, there are jobs appealing to parents with high-income aspirations, who in turn tend to focus on their children's educational attainment. Especially in countries where there is a focus on performance orientation, but also a high demand for good grades like in the ESEA countries, this might be related to more pressure on the shoulders of the teachers than in countries where there is less of an emphasis on these features and thus be associated with a lower sense of ITSE (c.f. Pang, 2011).

For private schools, a negative association appeared for the Anglo-Saxon (ES = -.10) and especially a strong negative effect for the Nordic (ES = -.93) cluster, but no significant association in the ESEA cluster. This is contrary to previous findings, which have shown a higher level of TSE for teachers in private schools (i.e. Fackler & Malmberg, 2016). The differences between the country clusters for private schools are more nuanced than for most of the other variables, as there is a bigger difference between the Anglo-Saxon and Nordic countries, which tend to share more similarities. The Anglo-Saxon countries are supposed to be more egalitarian and only support societal differences to a small degree, but so do the Nordic countries (House et al, 2004). One possible explanation might be that individualistic countries like the western ones are more open to individual choice and thus there is more acceptance of private schools, as they could be seen as individual decisions that support people's own choice rather than bigger societal goals. In the Anglo-Saxon countries, where in England, for instance, private schools have a long tradition, the expectation would have rather been that school type is strongly associated with ITSE, as it shapes the student body strongly due to the parents' socio-demographics (SES, education and income) (Eurydice, 2010). In the rather egalitarian and individualistic societies like the Nordic countries it might be possible that a private school is rather a rare thing and they may attract different types of students. Thus, their association with ITSE could be strongest due to their scarcity (Telhaug et al., 2006). Furthermore, in the Nordic countries teaching at a private school might let teachers face particular challenges, like parents who are highly educated themselves and send their children there with high expectations. Additionally, those parents may be more likely to question the teacher's expertise

and competence. Accordingly, what might seem to be a potential source of mastery experience from well-supported students, could turn into a setback due to pressure from more critical and challenging parents.

The last school characteristic investigated was student enrolment, which showed a too small to measure association with each cluster.

7.1.3 TO WHAT EXTENT IS TSE IN STUDENT ENGAGEMENT PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN THE THREE COUNTRY CLUSTERS?

The following section presents the results for separate multilevel models that predict teachers' self-efficacy in student engagement for each country cluster. It comprises a descriptive summary of the numeric results and at the end of the section, there is a comparative subsection that aims at explaining similarities and differences between the three different country clusters.

At first, a baseline model was calculated against which the full model was then compared. The -2log likelihood is the same for all three country clusters, as it refers to the tripartite structure of the model. Thus, it will only be reported in the beginning for all cluster-specific sub-sections, whereas all following parameters will be reported in the respective country cluster section. The -2log likelihood of the baseline model was 215,098.35 and decreased to 211,595.08 (Δ -3,503) after adding all explanatory variables at once.

Table 28 Results for TSE in student engagement for all three country clusters

	Anglo-Saxon Cuntries							Nordic Countries							East- & South-East Asian Countries						
	Baseline Model			Full model				Baseline Model			Full model				Baseline Model			Full model			
Response	β	p	S.E.	β	p	S.E.	ES	β	p	S.E.	β	p	S.E.	ES	β	p	S.E.	β	p	S.E.	ES
Fixed Part																					
intercept	3.187	***	.010	3.187	***	.010		3.187	***	.010	3.187	***	.010		3.187	***	.010	3.187	***	.010	
Teacher																					
Female Gender				.040	**	.014	.32				.003	**	.010	.03				.020	**	.010	.10
Educational Level				-.019	n.s.	.036	-.15				.018	n.s.	.014	.20				.012	n.s.	.017	.06
Teaching Exp.				.004	***	.001	.03				.003	***	.000	.03				.007	***	.001	.03
Hours Teaching				.004	***	.001	.03				.008	***	.001	.09				.003	***	.001	.01
Hours Planning				.003	**	.001	.02				.002	*	.001	.02				.001	n.s.	.001	.00
Hours Admin. Work				.005	**	.002	.04				.002	n.s.	.002	.02				.001	*	.001	.00
Construct. Bel.				.043	***	.004	.35				.050	***	.003	.55				.055	***	.033	.27
Classroom																					
Class size				.000	n.s.	.001	.00				.002	**	.001	.02				.000	n.s.	.001	.00
Classroom Climate				.050	***	.004	.40				.038	***	.003	.42				.033	***	.002	.16
Low Student Achievement				.007	n.s.	.008	.06				.010	n.s.	.007	.11				-.015	**	.006	-.07
Teaching & Learning				.001	n.s.	.002	.01				.001	n.s.	.001	.01				.002	n.s.	.001	.01
Keeping Order				-.002	n.s.	.002	-.02				.001	n.s.	.002	.01				.001	n.s.	.001	.00
Admin. Tasks				-.004	*	.002	-.03				-.001	n.s.	.002	-.01				.004	**	.001	.02
Different Language				.018	**	.007	.15				.009	n.s.	.007	.10				.029	***	.006	.14
Disadv. Background				.033	***	.008	.27				.030	***	.007	.33				.053	***	.007	.26
Principal																					
Female Gender				.026	n.s.	.019	.21				-.010	n.s.	.021	-.11				.143	***	.028	.71
Age				-.001	n.s.	.001	-.01				.001	n.s.	.001	.01				-.017	***	.002	-.08
Educational Level				-.019	n.s.	.031	-.15				-.022	n.s.	.035	-.24				.023	n.s.	.069	.11
Work Experience				-.002	n.s.	.002	-.02				.004	**	.002	.04				.009	**	.003	.04
Instructional Leadership				.006	n.s.	.005	.05				-.002	n.s.	.005	-.02				.092	***	.007	.46
School																					
Student Enrolment				.000	*	.000	.00				-.005	n.s.	.008	-.06				.000	***	.000	.00
Type of School				-.027	**	.020	-.22				-.197	***	.028	####				-.076	n.s.	.048	-.38
Commuity Size				.027	***	.008	.22				-.031	***	.000	-.34				-.064	***	.010	-.32
Random Part																					
School	.005	***	.002	.008	***	.002		.014	***	.002	.022	***	.002		.020	***	.025	.074	***	.005	
Teacher	.259	***	.006	.206	***	.006		.266	***	.005	.156	***	.004		.213	***	.004	.190	***	.004	
-2*loglikelihood	215098.35			211595.08				215098.35			211595.08				215098.35			211595.08			
Δ -2*loglikelihood				-3,503							-3,503							-3,503			
% variance				Explained variance				% variance			Explained variance				% variance			Explained variance			
School	1.89%			School	49.34%			School	5.00%		School	59.55%			School	8.58%		School	69.38%		
Teacher	98.11%			Teacher	n.s.			Teacher	95.00%		Teacher	n.s.			Teacher	91.42%		Teacher	n.s.		
ICC																					
School	1.89%	***	.010	3.74%	***	.009		5.00%	***	.015	12.36%	***	.013		8.58%	***	.015	28.03%	***	.015	
Teacher	98.11%	***	.010	96.26%	***	.009		95.00%	***	.015	87.64%	***	.013		91.42%	***	.015	71.97%	***	.015	

7.1.3.1 Anglo-Saxon countries

In the Anglo-Saxon country cluster, initially 1.89% of the unexplained variance resided among schools and 98.11% among teachers. By adding explanatory variables, among schools 49.34% more of the variance could be explained significantly, but no more among teachers.

All but one of the teacher characteristics (educational level) were significantly related to TSE in student engagement in Anglo-Saxon countries. Female gender (β = .040, p < .01, ES = .32) had a medium positive relationship to ETSE as well as constructivist beliefs about teaching (β = .043, p < .001, ES = .35). All other teacher related characteristics such as the years of teaching experience (β = .004, p < .001, ES = .03), weekly hours spent on teaching (β = .004, p < .001, ES = .03), spent on planning lessons (β = .003, p < .01, ES = .02) and on administrative work (β = .005, p < .01, ES = .04) only showed a weak positive association with ETSE.

From the eight suggested classroom characteristics, only half showed a significant association with ETSE. Classroom climate (β = .050, p < .001, ES = .40) was moderately positively related to ETSE, as well as the percentage of students with a different mother tongue than the language of instruction (β = .018, p < .01, ES = .15) or the percentage of students in the class from a disadvantaged socio-economic background (β = .033, p < .001, ES = .27). In contrast, out of the three suggested tasks a teacher has to perform in the classroom, only the percentage of time spent on administrative duties (β = -.004, p < .01, ES = -.03) had a weak negative relationship with ETSE.

Out of the five suggested principal characteristics, which were female gender, age, educational level, years of work experience as principal and an instructional style of leadership, none was significantly related to ETSE in the Anglo-Saxon countries.

In contrast to the principal characteristics, all three school characteristics showed a significant association with ETSE: student enrolment showed almost no relationship to ETSE (β = .000, p < .05, ES = .000), being a teacher at a private school, in contrast, was weakly negatively (β = -.027, p < .01, ES = -.22) and the community size was weakly positively (β = .027, p < .001, ES = .22) associated.

7.1.3.2 Nordic countries

For the Nordic countries, 5.00% of the initially unexplained variance resided among schools and 95.00% among teachers. After adding teacher, classroom, principal and school characteristics to the model, no more variance could significantly be explained among teachers, but 59.55% among schools.

From the suggested teacher characteristics, all but two were significantly related to ETSE in the Nordic countries: female gender (β = .003, p < .01, ES = .03), years of teaching experience (β = .003, p < .001, ES = .03), hours per week spent on teaching (β = .008, p < .001, ES = .09) and hours per week spent on planning lessons (β = .002, p < .05, ES = .02) were all very weakly positively related to ETSE. The only characteristic with a moderately strong association with ETSE were the constructivist beliefs about teaching a teacher has (β = .050, p < .001, ES = .55).

Out of the eight classroom characteristics, only three appeared to be significantly related to ETSE: the class size (β = .002, p < .01, ES = .02), a good classroom climate (β = .038, p < .001, ES = .42) and the percentage of students from a disadvantaged socio-economic background (β = .030, p < .001, ES = .33).

From the principal characteristics, only the work experience showed a significantly positive association with ETSE (β = .004, p < .01, ES = .04).

Investigating school characteristics and their relation to ETSE in Nordic countries showed that teachers in private schools have quite a lower sense of self-efficacy (β = -.197, p < .001, ES =-.22) compared to their colleagues teaching at public schools. In contrast, teachers working in schools in a bigger community had a somewhat higher sense of ETSE than their peers in smaller communities (β = -.031, p < .001, ES = -.34).

7.1.3.3 East- and South-East Asian countries

For East- and South-East Asian countries, 8.58% of unexplained variance initially resided among schools and 91.42% among teachers. By adding explanatory variables in the full model, no more variance could significantly be explained among teachers, but 69.38% more among teachers.

Among the suggested teacher characteristics, all but two showed a significant positive association with ETSE: female gender was weakly related to a higher sense of ETSE (β = .020, p < .01, ES = .10) as well as constructivist beliefs about teaching (β = .055, p < .001, ES = .27).

The years of teaching experience (β = .007, p < .001, ES = .03), hours per week spent on teaching (β = .003, p < .001, ES = .01) and the time spent on administrative work (β = .001, p < .05, ES = .00) showed very negligible associations with ETSE.

Investigating classroom characteristics and their association with ETSE in East- and South-East Asian countries showed that a good classroom climate (β = .033, p < .001, ES = .16), the percentage of children with another mother tongue (β = .029, p < .001, ES = .14) as well as the percentage of pupils from a socio-economically disadvantaged home (β = .053, p < .001, ES = .26) have a weak to moderate association with ETSE. Low student achievement, in contrast, showed a weak but negative relationship with ETSE (β = -.015, p < .01, ES = -.07). The percentage of classroom time spent on administrative tasks, finally, had a very weak positive association with ETSE (β = .004, p < .01, ES = .02).

Contrary to the other two country clusters, where almost no principal characteristics appeared to be associated with ETSE, all but one (principal's educational level) showed a significant relationship with ETSE in the ESEA countries. Female gender was quite strongly associated with a higher sense of ETSE (β = .143, p < .001, ES = .71) as well as an instructional leadership style (β = .092, p < .001, ES = .46). Principal age, in contrast, had a very weak but negative relationship with ETSE (β = -.017, p < .001, ES = -.08) and the principal's work experience a very weak positive association (β = .009, p < .01, ES = .04).

Out of the three school characteristics, only community size (β = -.064, p < .001, ES = -.32) was of moderate size.

7.1.4 SIMILARITIES AND DIFFERENCES IN THE COUNTRY CLUSTERS FOR TEACHER'S SELF-EFFICACY IN STUDENT ENGAGEMENT (ETSE)

The following sub-section discusses the differences in the results for teacher, classroom, principal and school characteristics across the three country clusters teachers' self-efficacy in student engagement. As for ITSE, effect sizes will be reported as unbiased comparative measures.

7.1.4.1 Teacher characteristics

The teacher characteristics investigated were female vs. male gender, educational level, years of teaching experience, weekly hours spent on teaching, preparing lessons or administrative work and the degree to which a teacher holds constructivist beliefs about teaching.

For teacher's self-efficacy in student engagement, all three clusters showed a positive association with gender; the strongest was found in the Anglo-Saxon countries (ES = .32) compared to ES = .03 within the Nordic countries and ES = .10 within the ESEA countries. This is – as with ITSE – against the hypothesis that TSE is higher for male teachers as found by, for example, Cousins et al. (1996). However, this result is in line with the comparatively higher scores that Anglo-Saxon countries had for gender egalitarianism, which tend to support women in the workforce and value ambitious female employees (House et al., 2004). Similar, though tentative explanations can be derived for the Nordic countries, but for the ESEA cluster the effect size is too small to derive meaningful comparisons.

The educational level, in contrast, was not significant in either cluster.

Regarding years of work experience, all three clusters showed the same weak relationship and effect size (ES = .03). Work experience can be interpreted as a means of gaining mastery in developing successful strategies to engage students in the lessons. This source seems to be equally important in the ESEA countries, which are very teacher-focused (c.f. Hofstede, 2011), and the Anglo-Saxon or Nordic countries, where the lessons tend to be more student centred (ibid.). However, this is partly in line with previous research on TSE (i.e. Cousins et al., 1996; Fackler & Malmberg, 2016).

Comparing the three types of work allocation, the association for the hours per week spent on teaching was significantly weakly related to ETSE in all three clusters (ES = .01 – ES = .09). For hours spent on planning lessons, there was no significant relationship with the ESEA cluster, but the others (ES = .02). Finally, the hours spent on administrative duties, were not significantly related to the Nordic cluster, but showed a weak positive association with the Anglo-Saxon (ES = .04) and ESEA (ES = .00) cluster. The size of these ES is almost negligible and thus no comparisons will be made.

In contrast, constructivist beliefs about teaching showed a significant positive association with ETSE in all three clusters, with the biggest effect size in the Nordic cluster (ES = .55) and the weakest in the ESEA cluster (ES = .27). The features of a constructivist belief about learning that entail a very self-responsible and engaged view of the learner, fits well with the teachers' own perceptions of their abilities to engage students in the lessons and foster their own inquiry, as is part of the measure of ETSE used here. Moreover, constructivist beliefs about teaching are associated with the teachers' taking on the responsibility for their students' learning, which can be seen an integral part of engaging them in the lessons: teachers, who see themselves as parts of their students' learning success, might well try to make them part of the lessons. This is in line with both the teaching approaches found in the Nordic and Anglo-Saxon countries (Klette et al., 2018; Telhaug et al., 2006; Mujis et al., 2018; Lindorff & Sammons, 2018). It seems that in more student-focused classroom settings as in the Nordic countries, constructivist beliefs contribute more to explaining variation in ETSE than in the more teacher-led ESEA classrooms.

7.1.4.2 Classroom characteristics

In total, eight classroom characteristics were tested as predictors in each country cluster. These were the class size, classroom climate, percentage of low achievers in the class, percentage of classroom time spent on teaching and learning, keeping order or administrative duties, the percentage of pupils with a different mother tongue than the language of instruction and from a disadvantaged home.

Regarding the class size, the only significant association was found with the Nordic cluster, but the effect is very small (ES = .02). Given the very small effect size it is not appropriate to

speculate on the differences. Rather it can be concluded that class size does not seem to be an important predictor of TSE in any country cluster.

The classroom climate revealed a positive association with ETSE in all three clusters, with a similar strengths of effect sizes in the Anglo-Saxon (ES = .40) and the Nordic (ES = .42) cluster but was notably weaker in the ESEA cluster (ES = .16). The classroom climate as defined here covers students that take care of a good classroom climate themselves, which, in turn, can provide a source of mastery experience as well as a source of feedback, if the students take action to find themselves in a supportive classroom environment. Furthermore, teachers with well engaged students might not have to beg for attention, as the students are focused on the lessons anyway. The stricter classroom management that has been observed in the ESEA countries (Pang, 2011) could hint that an engaging classroom climate plays a less important role for TSE in these countries than in the western ones, where teachers tend to face difficulties to enforce discipline in the classroom (Hargreaves, 2009). It is somewhat surprising that the classroom climate showed a stronger effect size in the two western clusters than in the eastern cluster. However, Hofstede (1986) found that students in classrooms with flat hierarchies are expected to be more independent and responsible themselves, which is supported by this measure of classroom climate. Alternatively, it could be assumed that teachers in the stricter ESEA countries have to worry less about student misbehaviour and thus the classroom climate plays a less important role there.

The percentage of low achievers in the classroom revealed only a significant negative association with ETSE in the ESEA cluster (ES = -.07). The association for the Anglo-Saxon and Nordic teachers was not significant. A possible explanation is that the high pressure put on teachers in the ESEA countries leads to a lower of ETSE in combination with low student achievement, which could result from disengaged students. However, the effect size is too small to draw sound conclusions.

Comparing the three types of time allocation shows that the time spent on teaching and learning or keeping order is not significantly related to ETSE in either cluster and the time spent on administrative tasks is only very weakly negatively associated in the Anglo-Saxon (ES = -.03) and ESEA (ES = .02) countries. However, the effect sizes are very tiny.

Investigating the relationship for the percentage of students in the classroom with a different mother tongue than the language of instruction showed a significant relationship between ETSE and the ESEA and Anglo-Saxon countries with similarly big effect sizes (ES = .14 and ES = .15). The percentage of students from a disadvantaged background, in contrast, showed a significant relationship with all three clusters and the effect sizes ranged from ES = .26 in the ESEA cluster to ES = .33 in the Nordic cluster. As already argued previously, these can be seen as contextual measures of the classroom composition. Hence, although, at first, teaching and engaging students might be more difficult in these settings, there is more room for having success, as the expectations of the teachers tend to be adjusted (c.f. Speybroeck et al., 2012) to the perceived student ability and thus mastery experience is achieved more easily. Perceived obstacles might be that teachers find it more difficult to get through to students who have a different mother tongue than the language of instruction or face more behaviour difficulties (Woessmann & West, 2010). However, looking at the migration policies of the ESEA countries in the cluster reveals that those are rather restrictive and mostly let in skilled workers, who can contribute to the economy of the countries (United Nations, 2017) and tend to have an ambitious attitude towards education. Accordingly, having a different mother tongue in those countries might not necessarily be associated with a low social status, but those children, in contrast, might come from backgrounds where education is appreciated so that migration is not necessarily a setback. Moving on to the percentage of students from a low SES and their positive association with ETSE: this could be explained by the ambitious aims Asian parents have for their children (Pang, 2011). Education is seen as a main means for success in later life and educational achievement is key to moving up in society, so that it can be assumed that parents from a low SES still tend to focus on their child's learning and those children are well engaged in the lessons. This is not the case to the same degree in the western countries, where the responsibility for student success is often rather attributed to the school system.

Finally, low student achievement only appeared to be significantly negatively associated with ETSE in the ESEA countries, but the effect size (ES = -.07) is negligibly small.

7.1.4.3 Principal characteristics

Five principal characteristics were tested in the models: Female vs. male gender, age, the educational level, years of work experience and the degree to which the principal employs an instructional leadership style.

For both, the principal's female gender and age, the only significant association with ETSE was found in the East and South-East Asian countries (ES = .71). The principal's work experience appeared to be significant in the Nordic and ESEA countries, but only very weakly (ES = .04 both times). The principal's instructional leadership style, again, only was significant in the ESEA countries (ES = .46). Generally speaking, it can be said that the principal seems to make more of a difference in predicting ETSE in the ESEA countries than in the two western country clusters as in the Anglo-Saxon cluster no and in the Nordic cluster only one predictor was significant for explaining variation in ETSE. As reported by Eagly et al. (1992), female principals tend to employ a more humane and democratic leadership style, which could be more important to explain variation in ETSE in the ESEA cluster. This has been found in previous research by for example Coladarci et al., (1992); Raudenbush et al., (2002) or Guo et al., (2012) and is in line with our hypothesis. It might possibly be that in a high-pressure environment like schools in the ESEA countries are described, these characteristics of leadership are particularly helpful. Moreover, given the comparatively high support for power distance in those countries, teachers might appreciate a principal who states clear academic goals that provide orientation. But so far, no studies exist that show a direct relationship between these two features, so the possible explanation is given rather tentatively.

Principal age, in contrast, had a very weak but negative relationship with ETSE (ES = -.08) and the principal's work experience a very weak positive association (ES = .04) In the ESEA cluster, but were not significant in the Anglo-Saxon and Nordic countries. Age and work experience can be seen of sources of mastery experience for the principals, as time allows them to grow into their role as principals. Relatively speaking, the ESEA countries scored most highly on the power distance scale used by House et al. (2004), which includes that superiors are perceived as important in organisations. However, the respective effect sizes are too small to draw strong conclusions.

7.1.4.4 School characteristics

Finally, there were three school characteristics tested in the model: Number of students enrolled at school, whether it is private or public school and the size of the community, in which the school is located. Investigating the differences for these school characteristics and the difference of their association with ETSE across the three different country clusters found multiple different results.

For private schools, a negative association with ETSE was found in the Anglo-Saxon (ES = -.22) and Nordic (ES = -.22) countries, but no association in the ESEA cluster. Again, the western countries are more similar to each other than either is with the eastern cluster. In the Nordic countries, private schools are relatively young, not very wide spread and have a mixed image (Telhaug et al., 2006); very much in contrast to the Anglo-Saxon countries, where private schools have long tradition (Eurydice, 2010). For the Anglo-Saxon countries it can be assumed that as private schools are often very carefully chosen by the parents and paid for, the composition of students might be different from the one at public schools (c.f. Yun & Reardon, 2005) and parental demands might be difficult to juggle. For the Nordic countries, almost contrariwise, is their focus on progressive and social-oriented schooling, which came along with a rejection of private schools (Telhaug et al., 2006). Provided these normative assumptions and the compositional factors, teachers in private schools might feel less self-efficacious in engaging their students than their colleagues in more widely respected public schools with a less selective studentship. Hence, due to the social standing of those schools the very strong negative association in the Nordic countries could be explored further. The notion conveyed in the literature is that schooling in the ESEA countries is highly regarded; regardless of its status as public or private and families tend to make huge financial investments in their children's education either by paying fees or for private tutoring. Thus, it might well be that the students in private as compared with public schools are not that different from each other to school than in private vs. public schools in the Anglo-Saxon or Nordic countries. Furthermore, the teachers in the ESEA countries have a high social standing nationally. However, the international literature on this is not very well developed and would benefit from more research.

Lastly, the school's location (in terms of community size) was found to be significantly related to ETSE in all three clusters but showed a negative relationship to ETSE in the ESEA (ES = -.32) and Nordic (ES = -.34), but a positive relationship with in the Anglo-Saxon (ES = .22) cluster. As already mentioned in the previous section, community size provides an indicator for bigger cities, which tend to attract a better educated work force due to the job market there; a rural-urban divide is often found. Thus, more eager students might be found in this environment, as there is a more positive attitude towards education among young people in this setting (Copus et al., 2006). Conversely, in bigger cities there also tends to be a more diverse population with potentially challenging students. With respect to the composition of the ESEA cluster it seems to be worth pointing out that with Singapore, there is one particularly big city in the cluster, which might shape the result in this direction. Overall, it is not possible from the findings to draw any conclusions about the positive association of community size with ETSE in the Anglo-Saxon and negative association in the other two country clusters.

7.1.5 TO WHAT EXTENT IS TSE IN CLASSROOM MANAGEMENT PREDICTED BY TEACHER, CLASSROOM, PRINCIPAL AND SCHOOL CHARACTERISTICS IN THE THREE COUNTRY CLUSTERS?

As with the previous models, the model fit for teachers' self-efficacy in classroom management will be reported in the introduction upfront, as it is the same for all three country clusters. All cluster specific information will then be reported in the respective results sections. In the following three section, only a brief description of the results will be given, and a comparative discussion follows in Section 7.4.6.

The initial -2log likelihood for TSE in classroom management was 192,537.17 and was reduced to 188,697.97 after the predictors were added in the full model (Δ_{-2LL} = -3,839).

7.1.5.1 Anglo-Saxon countries

Within the Anglo-Saxon countries, 3.66% of the initially unexplained variance resided among schools, 96.34% among teachers. By adding explanatory variables in the full model, no further variance could be explained among schools, but 1.22 % more among teachers.

All of the teacher characteristics showed a significant relationship with MTSE. Female gender (β = .072, p < .001, ES = .39) had a moderately strong positive association with MTSE, while

the teachers' educational level showed a moderately large negative relationship (β = -.090, p < .01, ES = -.49). The years of teaching experience (β = .003, p < .01, ES = .02), as well as the time per week spent on teaching (β = .003, p < .01, ES = .02) and time spent on administrative work (β = .008, p < .001, ES = .04) showed only a very weak positive association with MTSE, whereas the hours spent on planning lessons (β = -.004, p < .15, ES = -.02.) showed a very weak negative association with MTSE. The constructivist beliefs about teaching and learning a teacher has (β = .014, p < .001, ES = .08), showed a weak positive relation to the teachers' beliefs in their SE in classroom management.

Out of the eight suggested classroom characteristics, only four contributed to explaining variation in MTSE in Anglo-Saxon countries. A good classroom climate (β = .064, p < .001, ES = .35) as well as low student achievement (β = .034, p < .001, ES = .18) and the percentage of students from a disadvantaged home (β = .039, p < .001, ES = .21) had a weak or moderately strong positive association with MTSE. The percentage of time spent on administrative duties (β = -.004, p < .05, ES = -.02), in contrast, was only very weakly negatively related to MTSE.

Out of the five principal characteristics, none contributed to explaining variation in MTSE in teachers in the Anglo-Saxon countries.

Taking into account the school environment, only the number of students enrolled at the respective school (β = .000, p < .001, ES = .00) had very weak positive association with MTSE, but not whether it was a public or private school or the community size that mattered. The association between MTSE and the number of students enrolled, however, is too small to be displayed by the programme. For this reason, it will not be discussed here.

Table 29 Results for TSE in classroom management for all three country clusters

	Anglo-Saxon Cuntries							Nordic Countries							East- & South-East Asian Countries						
	Baseline Model			Full model				Baseline Model			Full model				Baseline Model			Full model			
Response	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**	**β**	**p**	**S.E.**	**β**	**p**	**S.E.**	**ES**
Fixed Part intercept	3.367	***	.019	3.367	***	.019		3.367	***	.019	3.367	***	.019		3.367	***	.019	3.367	***	.019	
Teacher																					
Female Gender				.072	***	.014	.39				.040	**	.012	.27				-.029	**	.011	-.17
Educational Level				-.090	**	.034	-.49				-.015	n.s.	.016	-.10				-.023	n.s.	.019	-.13
Teaching Exp.				.003	**	.001	.02				.005	***	.001	.03				.006	***	.001	.04
Hours Teaching				.003	**	.001	.02				.008	***	.001	.05				.002	**	.001	.01
Hours Planning				-.004	**	.001	-.02				-.002	n.s.	.001	-.01				-.002	*	.001	-.01
Hours Admin. Work				.008	***	.002	.04				.006	**	.002	.04				.004	***	.001	.02
Construct. Bel.				.014	***	.004	.08				.023	***	.003	.16				.049	***	.003	.29
Classroom																					
Class size				.000	n.s.	.001	.00				.003	***	.001	.02				.000	n.s.	.001	.00
Classroom Climate				.064	***	.004	.35				.061	***	.003	.42				.057	***	.003	.33
Low Student Achievement				.034	***	.009	.18				.039	***	.007	.27				.021	**	.007	.12
Teaching & Learning				-.001	n.s.	.002	-.01				.000	n.s.	.002	.00				.002	n.s.	.001	.01
Keeping Order				-.001	n.s.	.002	-.01				-.002	n.s.	.002	-.01				.003	**	.001	.02
Admin. Tasks				-.004	*	.002	-.02				-.001	n.s.	.002	-.01				.003	**	.002	.02
Different Language				-.001	n.s.	.007	-.01				-.006	n.s.	.007	-.04				.027	***	.006	.16
Disadv. Background				.039	***	.009	.21				.034	***	.008	.23				.053	***	.007	.31
Principal																					
Female Gender				-.028	n.s.	.018	-.15				-.020	n.s.	.015	-.14				.105	***	.023	.61
Age				-.002	n.s.	.001	-.01				.001	n.s.	.001	.01				-.011	***	.002	-.06
Educational Level				-.017	n.s.	.032	-.09				-.001	n.s.	.033	-.01				-.049	n.s.	.056	-.29
Work Experience				-.003	n.s.	.002	-.02				.003	**	.031	.02				.008	**	.003	.05
Instructional Leadership				.004	n.s.	.005	.02				.000	n.s.	.004	.00				.062	***	.005	.36
School																					
Student Enrolment				.000	***	.000	.00				.000	***	.000	.00				.000	***	.000	.00
Private School				-.017	n.s.	.019	-.09				-.130	***	.024	-.89				-.071	*	.040	-.41
Commuity Size				.005	n.s.	.008	.03				-.005	n.s.	.007	-.03				-.086	***	.008	-.50
Random Part																					
School	.007	***	.003	.006	***	.002		.005	***	.001	.009	***	.002		.275	***	.010	.042	***	.003	
Teacher	.184	***	.005	.236	***	.006		.187	***	.004	.231	***	.005		.169	***	.003	.271	***	.005	
-2*loglikelihood	192537.17			188697.97				192537.17			188697.97				192537.17			188697.97			
Δ -2*loglikelihood				-3,839							-3,839							-3,839			
% variance				Explained variance				% variance			Explained variance				% variance			Explained variance			
School	3.66%			School	n.s.			School	2.60%		School	30.56%			School	61.94%		School	n.s.		
Teacher	96.34%			Teacher	1.22%			Teacher	97.40%		Teacher	n.s.			Teacher	38.06%		Teacher	56.04%		
ICC																					
School	3.66%	***	.008	2.48%	**	.008		2.60%	***	.008	3.75%	***	.007		61.94%	***	.014	13.42%	**	.010	
Teacher	96.34%	***	.008	97.52%	***	.008		97.40%	***	.008	96.25%	***	.007		38.06%	***	.014	86.58%	***	.010	

7.1.5.2 Nordic Countries

Within the Nordic countries, 2.60% of the initially unexplained variance was found among schools and 97.40% among teachers. After adding the teacher, classroom, principal and school characteristics in the full model, no further variance could be explained among teachers, but 30.56% more among teachers.

Out of the seven suggested teacher characteristics, four showed a significant association with MTSE: female gender (β = .040, p < .01, ES = .27) and constructivist beliefs about teaching (β = .023, p < .001, ES = .16) had a small positive association. Teaching experience (β = .005, p < .001, ES = .03), time per week spent on teaching (β = .008, p < .001, ES = .05) and time per week spent on administrative duties (β = .006, p < .01, ES = .04), in contrast, only showed only a very weak positive association with MTSE.

Four of the eight classroom characteristics showed a positive association with MTSE in the Nordic countries. The class size (β = .003, p < .001, ES = .02) had a very weak positive relation. A moderately strong positive association was found for a good classroom climate (β = .061, p < .001, ES = .42), low student achievement (β = .039, p < .001, ES = .27) and a higher percentage of children from disadvantaged homes (β = .034, p < .001, ES = .23) in the classroom.

Only the principal's work experience as a principal (β = .003, p < .05, ES = .02) showed a positive but negligible relation to MTSE. All other principal characteristics appeared not to be significantly related to MTSE.

School characteristics that showed a negligible relation with the sense of MTSE for teachers in Nordic countries was the number of students enrolled at the school (β = .000, p < .001, ES = .00), but by contrast, teachers working in private schools (β = -.130, p < .001, ES = -.89) felt significantly less self-efficacious in managing the students' behaviour than their colleagues at public schools.

7.1.5.3 East- and South-East Asian countries

In the East- and South-East Asian countries, 61.94% of the initially unexplained variance was found among schools, 38.06% among teachers. This is notably the smallest amount of unexplained variance among schools found in any of the country clusters for any of the domains

of TSE. This points towards an interesting dynamic between East- and South-East Asian teachers and schools with regards to self-efficacy in managing students. By adding the teacher-, classroom-, principal- and school characteristics, no more variation among the schools, but 56.04% more among the teachers could be explained.

Adding the teacher characteristics revealed that all but the educational level were significantly related to MTSE in East- and South-East Asian countries. Female gender (β = -.029, p < .05, ES = -.17) showed a moderately weak but negative relation with MTSE, time spent planning lessons (β = -.002, p < .01, ES = -.01) a very weak negative relation to MTSE. The years of teaching experience (β = .006, p < .001, ES = .04), in contrast, had a very weak positive relation to MTSE, as did the hours per week spent on teaching (β = .002, p < .05, ES = .01) and the hours per week spent on administrative duties (β = .004, p < .001, ES = .02). The degree to which the teacher holds constructivist beliefs about teaching (β = .049, p < .001, ES = .29), then, showed a modest positive association with MTSE.

From the classroom predictors, all but the class size and the time in the classroom spent on teaching and learning were significantly associated with MTSE. A good classroom climate (β = .057, p < .001, ES = .33) showed a moderately strong positive association with MTSE, as did the percentage of pupils from a disadvantaged background (β = .053, p < .001, ES = .31) in the classroom. Low student achievement (β = .021, p < .01, ES = .12) and the percentage of students whose mother tongue was different from the language of instruction (β = .027, p < .001, ES = .16), then, showed a small positive association with MTSE. Finally, the percentage of time spent on keeping order (β = .003, p < .01, ES = .02) and spent on administrative tasks (β = .003, p < .01, ES = .02) showed negligible associations with MTSE in the East- and South-East Asian countries. It is contrary to my hypothesis that the time spent on keeping order in the classroom is positively related to MTSE.

A different picture emerges for the principal characteristics, where all but the principal's educational level showed a significant association with the teachers' sense of self-efficacy in classroom management. Female gender (β = .105, p < .001, ES = .61) showed a strong positive association. There was no hypothesis guiding this research question, so this result is explorative. The degree to which the principal employs an Instructional leadership style (β = .062, p < .001, ES = .36) had a moderately strong positive relation to MTSE but the principal's work

experience (**β** = .008, p < .01, ES = .05) a very weak positive association. For the instructional leadership style, it can be assumed similar to the other aspects TSE, that the role model behaviour and goal setting of the principal, which is entailed in this leadership style, provides an environment in which the teachers' self-efficacy can develop. The principal's age (**β** = -.011, p < .001, ES = -.06), in contrast, showed a very weak negative relation to MTSE.

From the three school characteristics, all were significantly related to the teachers' belief in their classroom management efficacy. Student enrolment (**β** = .000, p < .001, ES = .00) showed a very weak positive association, whereas the status as a private school (**β** = -.071, p < .05, ES = -.41) and the community size (**β** = -.086, p < .001, ES = -.50) a moderately strong negative association with MTSE.

7.1.6 Similarities and differences in the country clusters for teacher's self-efficacy in Classroom Management (MTSE)

This sub-section investigates the differences that occur between the country clusters with respect to the association that MTSE has with each of the teacher-, classroom-, principal- and school characteristics. The measure used here will be the effect size as in the previous sections, as it is independent of the sample size and thus provides an unbiased measure to compare the three clusters.

7.1.6.1 Teacher characteristics

The teacher characteristics tested as possible predictors in the models were female gender, educational level, years of teaching experience, weekly hours spent on teaching, planning lessons or administrative tasks and the degree to which constructivist beliefs about teaching were held.

Investigating the difference for teacher gender, revealed differences between all three country clusters: although the effect size was much stronger in the Anglo-Saxon countries (ES = .39) compared to the Nordic countries (ES = .27), both were positive. In the ESEA countries, in contrast, the association was weaker but negative (ES = -.17), which is in line with my hypothesis that male teachers feel more self-efficacious (i.e. Cousins et al., 1992). The ESEA countries scored the lowest in terms of gender egalitarianism among the three clusters.

Hence, female teachers might face more difficulties managing difficult students or maintaining classroom discipline than their female colleagues in the western countries, which scored higher for gender egalitarianism and thus revealed greater acceptance of female authority (House et al., 2004).

For the teacher's educational level, the only significant negative association was found in the Anglo-Saxon countries (ES = -.49), whereas the association with MTSE was not significant in the other two clusters. Here, the relationship is different as for self-efficacy in instruction, where a higher level of SE was associated with a higher educational degree. A possible explanation might be found in John Dewey's idea of democratic education, which puts the teacher in the role of a 'supervisor of the class' and could undermine his authority in comparison to the traditional role of the teacher. Furthermore, it might be that more highly educated teachers experience a lack of authority, as they might not be able to bring across the content they would like to teach due to a lack of classroom structure (Norris, 2004). This could mean that after their formal education, teachers feel more prepared for the content knowledge they are supposed to deliver than for managing students and managing classroom discipline. This is in line with previous research, which has shown that teacher training courses do not contain enough pedagogic elements to make the teachers feel well prepared to manage their students' behaviour (König et al., 2011).

For the teacher's work experience, the effect size was similarly weak in all three country clusters (ES = .02 – ES = .04). Hence it might not be valid to draw sound conclusions to explain differences here. However, previous research suggested that work experience adds to a teacher's abilities to manage classrooms well, as they learn to interpret the students' signals and be better prepared to respond adequately (Gkolia et al., 2014; Zee & Koomen, 2016).

The picture for the work allocation per week was found to be rather heterogenic. The difference in the hours per week spent on teaching ranged from ES = .01 in the ESEA countries to ES = .05 in the Nordic countries and was thus very small. The weekly hours spent on planning lessons, then, was only significantly but negatively related to MTSE in the ESEA (ES = -.01) and Anglo-Saxon (ES = -.02), but not the Nordic countries. For hours per week spent on administrative duties, the effect sizes were small and positive for all three country clusters (ES = .02

– ES = .04). Due to the very small effect sizes, those differences will only be interpreted cautiously. Generally, it can be said that the types of time allocation are associated with a degree to which teachers can perform their duties in the classroom. Thus, it might be that i.e. teachers in a less hierarchical society like in the Nordic countries feel less discouraged when their lesson plans are disrupted by noisy students than teachers in more hierarchical societies, where a lack of respect might be perceived more severely.

Finally, the associations for constructivist beliefs about teaching and their effect sizes for MTSE in the three country clusters were all significant, but different in magnitude: The smallest effect size with ES = .08 was found in the Anglo-Saxon countries, and the biggest in the ESEA countries with ES = .29. Given that the perceived pressure on the teachers in the ESEA countries is quite high (Sato & McLaughlin, 1992; Lim, 2007) compared to the other two clusters, it might be that teachers who are convinced that students are responsible for their own learning and thus remove some of the pressure from the shoulders are able to develop a more positive attitude towards their abilities in maintaining a beneficial classroom climate (Sato & McLaughlin, 1992). It seems to be noteworthy that constructivist beliefs about teaching contributed significantly to explaining variation in all dimensions of TSE in all clusters. Accordingly, they seem to be supportive for different types of TSE.

7.1.6.2 Classroom characteristics

Regarding the classroom, the class size, the perceived classroom climate and the percentage of low achievers were investigated. Furthermore, time allocation in the classroom as to how much time was spent on teaching and learning, keeping order or administrative tasks as well as the percentage of students with a different mother tongue and the percentage of students from a disadvantaged home were analysed.

Investigating the association with MTSE and the class size only found a measurable relationship for the Nordic countries (ES = .02), but the effect size is negligibly small.

With respect to the classroom climate and the magnitude of its association with MTSE, all three country clusters had a moderately strong positive effect size (ES = .33 in the ESEA countries to ES = .42 in the Nordic countries). The positive association between MTSE and classroom climate is in line with the findings for ITSE and ETSE, but also previous research (i.e.

Ashton et al. 1983). Part of this concept of a good classroom climate is that the students themselves take care of it, so the teachers might feel supported in their efforts to create a stimulating learning environment (Künsting et al., 2016). Thus, they experience a source of mastery experience, but also verbal persuasion, as they receive student feedback. Furthermore, Rimm-Kaufmann & Sawyer (2004) found that high TSE teachers tend to apply stricter classroom management methods and thus are more likely to achieve a good classroom climate, if they are not coercive. Especially in the ESEA countries, where, for example, for Japan it is reported that the teacher is to be treated with respect (Sato & McLaughlin, 1992), bad classroom climate could be interpreted as disrespectful behaviour. Furthermore, the predominant way of teaching used is whole-class-teaching (Stevenson & Lee, 1995), which requires attentive students and seems to make successful classroom management even more crucial. But, also in the Anglo-Saxon, classroom discipline can be regarded important; especially as teachers tend to receive low levels of unconditional respect form their students (Lewis et al., 2005). Thus, the application of successful classroom management strategies might provide them with a particular sense of achievement. The positive association in all three clusters is not surprising, though, as it can be seen as a direct feedback loop and source of mastery experience (Bandura, 1986) of the teacher's classroom management skills, which should be valid across countries, regardless of cultural norms.

The association for the percentage of low achievers in the classroom was significantly positively related to MTSE in all three clusters. The strongest association was found in the Nordic countries (ES = .27), the weakest for the ESEA countries (ES = .12) and is in all three cases against my hypothesis that high student achievement should be associated with a high level of TSE (i.e. Mojavezi & Tamiz, 2012; Alivernini & Lucidi, 2011; Caprara et al., 2006). Here, regardless of cultural background and different values supported in the countries, it may be that teachers in all three country clusters seem to adjust their expectations towards the students' abilities (Speybroeck et al., 2012) and this, in turn, is associated with MTSE. However, similar explanations as for ITSE and ETSE might be possible here.

The time spent on different classroom activities drew a heterogenic picture, but the effect sizes were negligibly small: the percentage of time spent on administrative duties showed a significant positive association with MTSE in the Anglo-Saxon countries (ES = .02) and a small

negative association in the ESEA cluster (ES = -.02), but not so in the Nordic cluster. The percentage of time spent on keeping order in the classroom only showed a significant positive association in ESEA countries (ES = .02), but again no significant relationship in the other two clusters. The time spent on teaching and learning was not significantly associated with MTSE in either cluster. This result is particularly surprising, as according to Bandura's theory (1986) the time spent on teaching and learning could be interpreted as a result of successful classroom management, but no empirical investigations on this are available, to date.

For the percentage of students in the class whose mother tongue is different from the language of instruction only showed a significant positive association with MTSE in the ESEA countries (ES = .16), but not the other two clusters. A possible explanation here might be found in the social structure in the different country clusters. The ESEA countries tend to have a lesser influx of migrants and are focused on a few languages (i.e. Chinese and English), which might result in more homogeneous classrooms, which were found to be supportive of a higher level of MTSE (Geerlings et al., 2018), whereas, for example, in the Anglo-Saxon cluster, migration and thus students speaking a different mother tongue have long been there and thus teachers might be more used to behaviour management in diverse settings.

The percentage of students from socio-economically disadvantaged backgrounds showed a positive and significant relationship in all three country clusters; the strongest, though, appeared in the ESEA countries (ES = .31), the weakest in the Anglo-Saxon countries (ES = .21). The result that a low SES, which is often found to be related to difficult behaviour (White, 2010), is positively related to TSE, is surprising. However, in all three clusters, again expectation management might play a role (c.f. Speybroek et al., 2012). Thus, two explanations are possible: On the one hand, teachers could adjust their expectations to the students' perceived level of discipline, on the other hand, the parents' higher aspirations are reflected in the students' behaviour. From the current state of the art in research on cultural values, it is not possible to point out tentative explanations with respect to cultural differences.

7.1.6.3 Principal characteristics

Concerning the principals, their gender, age, educational level, years of work experience as principals and degree of instructional leadership style were analysed.

The principal's female gender only showed a significant positive relationship with MTSE in the ESEA countries and the ES = .61 was strong. Eagle et al. (1992), however, found in a meta-analysis that female principals tend to apply more reciprocal and humane leadership styles, which is often associated with a more open and also pleasant work environment. In countries that have a larger degree of power distance, as is the case in the ESEA countries (House et al., 2004; Hofstede, 2011), this might bridge the gap between teachers and principals and provide fruitful ways of manging difficult students. Hence, displaying empathy in leadership may be a generally beneficial approach to foster TSE. Given the association that also became apparent for the other two dimensions of TSE, female principals seem to be beneficial for all types of TSE investigated. For the principal's age, again, the only significant association was found in the ESEA countries, but negative and the effect size (ES = -.06) was very small.

The educational level of the principal had no significant relationship with MTSE in all three clusters.

For the principal's work experience, a significant relationship with MTSE was found for the ESEA cluster (ES = .05) as well as the Nordic countries (ES = .02). Both times, the association and effect size were negligibly small.

With respect to the instructional leadership style, only the ESEA countries had a significant and moderate positive relationship with MTSE (ES = .36), but not the other two country clusters. Power distance might serve as a possible source of explanation here (Hofstede, 1986; House et al., 2004). A principal who is giving direct guidance and is stating clear goals may provide orientation and serve as a role model. This might be less the case in countries that support rather flat hierarchies and teachers are given more freedom generally, as in the Anglo-Saxon and Nordic clusters. Furthermore, it can be assumed similar to the other aspects of TSE, that the role model behaviour and goal setting of the principal, which is entailed in this leadership style, provide an environment in which the teachers' self-efficacy in classroom management can develop (Bandura, 1977; Guskey & Passaro, 1994; Walker & Slear, 2011).

7.1.6.4 School characteristics

Finally, three school characteristics were compared: The number of students enrolled at school, the status as a private or public school and the size of the community in which the

school is located. For those school characteristics, the associations with MTSE in the country clusters were again very heterogenic.

The community size was only significantly related to MTSE in the ESEA countries (ES = -.50). The effect size is moderately strong indicating that a more diverse student body, which tends to be found in bigger cities, are more strongly negatively related to MTSE in ESEA than in the two western country clusters. Regarding the status of the school as 'private' showed a negative association in the ESEA (ES = -.41) and the Nordic countries (ES = -.89), but not the Anglo-Saxon cluster. The effect size in the Nordic countries is particularly noteworthy, as it is the strongest for any of the predictors with MTSE. Interestingly, the challenges that teachers may be experiencing in private schools (i.e. more pressure from the parents) are comparatively strongly related to MTSE in a negative way, although the expectation could have been that students from more affluent backgrounds display less problematic behaviour (i.e. Woessmann & West, 2010). Pinnock (2006) and Caprara et al. (2003) found that a higher level of TSE was associated with good parent-teacher relationships, which could explain the negative association with the apparently more challenging environment in private schools. However, it could also be argued that here the adjustment of the teachers' expectations for student behaviour work the other way around as in the classroom (Speybroeck et al., 2012): if the students fail the higher expectations of the teacher, the relative interpretation of bad behaviour might be worse than in public schools. Here, culture does not seem to mediate the teacher's perceptions.

Finally, the relationship between MTSE and student enrolment was too small in each cluster to produce meaningful results.

7.2 SUMMARY

The three sections above were investigating the differences between Anglo-Saxon, Nordic and East and South-East Asian countries in terms of a group of predictors (teacher, classroom, principal and school) of teachers' self-efficacy in instruction, student engagement and classroom management.

Although there are some striking differences both across clusters and dimensions of TSE, there are also a number of similarities in the findings across the countries; in particular with regard to characteristics that concern the teachers themselves. It might be that schools are in reality rather similar across countries due to characteristics they share regardless of wider culture or political systems. Those characteristics include the teaching and education of young people and their preparation for life and work in an institutionalised context. With this, similar demands might be placed on the teachers across countries. Moreover, classrooms tend to be set-up in a similar way and thus teachers work in similar environments across the globe. This might explain, why the more the characteristics concern the direct environment, a, for example the classroom setting, the less variation is found both across the country clusters and the different types of TSE. Consequently, when the wider environment is taken into account as determined by the principal and the school, the more variation there is and the more differently these predictors are associated across countries and dimensions of TSE.

7.2.1 DIFFERENCES IN THE EXPLANATION OF VARIATION ACROSS LEVELS

First of all, the ESEA countries stand out with respect to the distribution of the unexplained variance between the two levels for MTSE. Whereas the bulk of unexplained variance is found among teachers and only little among schools in the other country clusters, here it is the other way around: 38.08% resides among teachers, 61.94% among principals. By adding explanatory variables 56.04% of the variance can be explained among teachers, but this is still considerably less than within the other clusters, where about 95% of the variation among teachers could be explained. Hence it seems that East- and South-East Asian teachers are less different with respect to their sense of self-efficacy in classroom management than the other countries, but schools are making a bigger difference.

7.2.2 DIFFERENCES IN THE ASSOCIATION OF THE PREDICTORS

Teacher characteristics were in sum more important for explaining variation in TSE both across the three country clusters, and also across the three dimensions of TSE. Only with respect to the teachers' educational level and aspects of the teachers' workload (i.e. the time spent on planning lessons) differences – although only small ones - were found.

Within the classroom predictors, there is more variation than within the teacher predictors, but also quite a large proportion of them seemed to be relevant for explaining variation in all three country clusters and all three dimensions of TSE. One major difference was found for the Nordic countries, where the fewest number of predictors was significant; particularly types of time allocation in the classroom and partly student SES for TSE in student engagement or classroom management. Most classroom predictors, in contrast, were relevant for explaining variation in all domains of TSE in the ESEA countries. Finally, the Anglo-Saxon countries showed the biggest variation with respect to associations of classroom predictors across the three domains of TSE: student achievement, for instance, contributed significantly to explaining variation in MTSE, but not the other two domains, whereas the students' mother tongue only had a significant relationship with ETSE. Thus, it can be argued that the Anglo-Saxon countries show the greatest sensitivity in the three domains with regards to classroom compositional factors.

The biggest difference observed for the principal characteristics are found between the Anglo-Saxon countries where none of them was significantly contributing to explaining variation in either dimension of TSE and the ESEA countries, whereas all of them except the principal's educational level were contributing to explaining differences. In the Nordic countries, only work experience was significant for explaining variation in ETSE and MTSE and the educational level and instructional leadership style for ITSE. A key finding here is that between the two western clusters, there are more similarities with regard to the principal traits. This might be due to the egalitarian structures the countries in the clusters support.

Finally, for the school characteristics, the picture was more homogeneous: all of them with little variation across countries and dimensions became significant. However, differences between the countries were found for all three suggested predictors; student enrolment, type of school and community size. So, for example, all three predictors contributed to explaining variation in the Anglo-Saxon countries for ITSE and ETSE, which was the case for the ESEA countries and MTSE. In the Nordic countries, the picture was more diverse across the three dimensions of TSE. However, it seems to be noteworthy that the school environment seems to make more of a difference in explaining variation in TSE than the principal.

CHAPTER 8 DISCUSSION AND CONCLUSIONS

This chapter brings together the main findings of the first and second research question and discusses their links with the results of the theoretical framework and review of existing literature on TSE presented in Chapter 2. The study has used secondary analysis and a range of statistical models to pursue its aims. Firstly, the study sought to investigate predictors in relation to teacher-, classroom-, principal- and school characteristics of teachers' self-efficacy in the three TSE domains instruction, student engagement and classroom management across 32 OECD countries (1st RQ). Furthermore, it had the aim to compare those predictors across three country clusters representing Anglo-Saxon, Nordic and East and South-East Asian countries for the same dimensions of TSE (2nd RQ). These analyses sought to explore the different environments in which teachers work and associations with their own understanding of their efficacy as a teacher.

In the following sections of this chapter, the main results and contributions to the field of research on TSE are highlighted and the limitations and challenges of the study will be discussed. Implications for future research and education policy will conclude the chapter.

8.1 MAIN RESULTS AND CONTRIBUTIONS TO THE FIELD OF RESEARCH ON TEACHERS' SELF-EFFICACY

As discussed in Chapter 2, self-efficacy theory proposes four possible sources that are relevant for a person to develop the conviction that they have the necessary skills to fulfil a task. Those are first and foremost mastery experience or success; vicarious experience or a successful role model; verbal persuasion or positive feedback and finally emotional arousal or positive feelings associated with a task. In past research on teachers' self-efficacy, a tripartite instrument covering three domains has reached agreement and is now widely used: self-efficacy in instruction, student engagement and classroom management, which are supposed to represent the structure of frequently performed tasks in the classroom.

The data for the present study was taken from the Teaching And Learning International Survey (TALIS) published by the OECD in 2013, covering 32 countries with about 6.500 schools and 104.000 teachers and the sub-sample of Anglo-Saxon-, Nordic- and East- and South-East

Asian countries respectively. The country clusters were chosen according to a theoretical model proposed by Bulle (2011), who grouped countries according to broad features of their education systems like a progressive versus an academic trend and similarities in the structure of the curriculum. According to Bulle (2011) those features supposedly shape the environment in which the teachers work and where their sense of self-efficacy develops.

The next sub-section will highlight the contributions of this study to knowledge on teacher's self-efficacy and how they are linked to past research.

8.1.1 RESEARCH QUESTION ONE: TEACHER'S SELF-EFFICACY IN AN INTERNATIONAL CONTEXT

The objective of the first research question was to identify predictors of teacher's self-efficacy in the domains of student engagement, instruction and classroom management across a data-set comprising 32 OECD countries. This approach was chosen for the following reasons: Firstly, in previous studies a lack of international research on TSE that investigates other than American teachers has been identified (c.f. chapter 2 and Klassen et al., 2010). This might be because the concept was developed in America and thus the natural approach was to first apply it there until it reached international recognition. Hence, those countries have a 'head start' in research on TSE. In a globalising world where research is aiming to become internationally available, those concepts need to be tested if their application is possible beyond the original context. The TALIS Survey included measures of TSE demonstrating increasing global interest in the concept but, as yet, these data have not been exploited to explore the two main RQs identified for this thesis. The availability of data from 32 OECD countries has enabled secondary analyses to provide an overview of the application of TSE measures in an international context and to establish whether results previously based on studies of mainly American teachers can be replicated internationally. As can be seen in Chapter 4, this was possible. For instance, that a higher level of TSE was related to constructivist beliefs about teaching, a good classroom climate or the principal's instructional leadership style, were identified as significant predictors across 32 countries, which aligns with findings from previous studies.

Secondly, there are only few representative samples, which allow to draw generalisable conclusions. The OECD was able to collect data from 32 countries with an appropriately

rigorous and complex research design (allowing for clustering in the data at teacher, school and country level) and close monitoring of the data collection process. By utilising these TALIS data for secondary analysis, it was possible to contribute to the literature on TSE by using appropriate multilevel statistical models to address the main RQs and so produce robust and generalisable results.

Thirdly, in previous research most studies tend to focus on the teacher- and classroom factors but disregard the possible influence of the principal and wider school environment. In this study, for instance, only 12 articles were found to be relevant for describing the school level predictors, whereas about 30 references were made to the teacher level (c.f. section 2.3 ff). By combining information from the TALIS teacher- and principal questionnaire, the perspective of possible predictors was broadened, and light was shed on the potential role the principal and school environment may play in providing impact for a high sense of a teacher's self-efficacy. Nonetheless the statistical findings suggest that the direct work environment like the classroom and the persona of the teacher are more important in predicting variations in different domains of TSE than wider structures. This conclusion is supported by the result that most variation was found between teachers and least between schools.

Fourthly, the tripartite measure of teacher's self-efficacy in instruction, classroom management and student engagement has hardly been applied internationally, although it is the measure of TSE which receives most agreement in the literature. Only in recent years, as demonstrated by Pulou, (2007); Klassen et al., (2013); Gkolia et al., (2014); Künsting et al., (2016); Geerlings et al., (2018), the instrument has received more international recognition. However, there are measures that were developed in a certain national context (i.e. the Norwegian TSE instrument; Skaalvik & Skaalvik, 2007). Nonetheless, those approaches rather scatter the field (although it is valuable from the perspective of theoretical development), whereas this study was able to consolidate research on TSE by applying an established measure in many new geographical areas and demonstrating country invariance. Furthermore, the use of the tripartite measure proposed by Tschannen-Moran & Hoy (2001) in the more recent studies helped explain the results of this study in greater detail.

8.1.2 RESEARCH QUESTION TWO: COUNTRY CLUSTER COMPARISONS OF PREDICTORS OF THREE DOMAINS OF TEACHER'S SELF-EFFICACY

The objective of the second research question was to gain clearer understanding on how predictors of teacher's self-efficacy in student engagement, instruction and classroom management differ across a set of three country clusters comprising Anglo-Saxon, Nordic and East- and South-East Asian countries, which were chosen according to shared characteristics of the education systems (Bulle, 2011). With this it was hoped to gain new understanding on how the cultural environment and the school system, which are in this context seen as constituting the broader work context of teachers, may contribute to explaining differences in the teachers' perceptions for the three domains of TSE. It has to be borne in mind, however, that a higher sense of self-efficacy is always seen as desirable (Bandura, 1977 & 1996). The additional country cluster analyses that addressed RQ2 provided the opportunity to understand and learn from country differences and apart from contributing to research in the field of teacher's self-efficacy, to inform education policy and practice. The contributions to research on TSE of this second research question can be described as follows.

Firstly, building on the models developed for the first research question, which chose an overview perspective on predictors of TSE, this research question goes more in-depth by focusing on three different large country clusters. The three country clusters make it possible to gain insight into those by adding a cultural perspective, which considers values that constitute social relationships like the degree of power distance, collectivistic vs. individualistic attitudes or the support for gender egalitarianism, which can be seen as the philosophical underpinnings that shape the ideas of how a society should operate and the role of the education system in it. In the first research question, there was a fair amount of variation found on the country level (25-30%), which remained statistically unexplained because of a lack of predictors linked to the third (country) level. Hence, it offers a new perspective on teacher's self-efficacy with regards to the education systems the teachers work in and a cultural framework, which offer insight in the work conditions of teachers. The cultural norms supported in those with, for instance, respect to the perceived power distance between teachers and students or teachers and principals, or gender egalitarianism which leads to the respect of women in the workplace, offer insight into how those differences might be formed and help distinguish the clusters from each other. Hence, the results discussed in Chapter 7 add a comparative

cultural perspective to the past research on teacher's self-efficacy, which tends to focus on studies conducted in only a few cultures.

Secondly, the tripartite conceptualisation comprising measures for teacher's self-efficacy in student engagement, instruction and classroom management, although widely accepted, has still not yet been applied in many different cultural contexts beyond the US. For instance, German teachers have been studied by Künsting et al., (2016); Greek teachers by Poulou (2007) and Dutch teachers by Geerlings et al., (2007). However, there is still a gap in the international literature on the use of this tripartite measure of TSE by its use of data and findings related to teachers from 32 different countries. This provides a better understanding of the concept and the finding of country invariance for the three domains is of theoretical importance for the TSE field.

Thirdly, although the body of research on TSE over the past 40 years is vast, there is, as shown in the literature review (Chapter 2) – although thoroughly grounded in theory - little systematic research has been done in terms of the choice of predictors that have been analysed, which makes direct comparisons difficult. In this study, in addition to replicating the same models across three different country clusters, effect sizes were reported that provide a suitable and comparable measure of the magnitude of the association of the predictors with each of the three domains of TSE. Hence, this will help distinguishing predictors with more explanatory power from predictors with less, in the future.

8.2 STRENGTHS AND LIMITATIONS OF THE STUDY

As the previous section has shown, this study was able to contribute to advancing the field of research on teacher's self-efficacy in multiple ways. However, as with all research, in addition to the strength, there are also limitations to the study. These limitations will be discussed in the following section. As already mentioned above, the properties of the data set, the multilevel analysis and the country clustering advance the field of research on teacher's self-efficacy in different ways. First the strengths, then the limitations of the study will be discussed.

8.2.1 A COMPREHENSIVE SURVEY DESIGN

For the first research question, the full data set comprising 32 OECD and partner countries was used. Although this is a rather comprehensive data set, it has some selection bias to it. First of all, only those countries that can fulfil the criteria to be OECD members, which is to agree to work on their and their partners' economic growth (see https://www.oecd.org/legal/accession-process.htm), or are on the verge of it, were eligible to take part in the study. This leaves out countries that either do not fulfil the criteria or are not interested in a membership for various reasons. In addition, participation in the study was optional, which further limits the country sample. In that respect, the results are, strictly speaking, based on a selection of countries whose policy makers support the OECD's perspective of education as a means to promote economic outcomes. This is not a limitation of the quality of the data or the trustworthiness of the results but provides information about the motivation and means of the data collection. Nonetheless, the secondary analysis of these OECD data enabled the largest and most comprehensive study of the predictors of TSE across three domains so far for a wide range of countries that meet the OECD requirements.

The second research question, in contrast, is not affected by this bias to quite the same degree. The countries that form the three clusters were chosen according to theoretical assumptions about features of their education systems testing the theoretical model proposed by Bulle (2011). The data availability still depends on the TALIS data set (along with its recognised strengths and limitations) but thanks to data availability the clusters contain almost a complete set of countries belonging to them. During the interpretation of the results it became apparent that there are more similarities in findings between the Nordic and Anglo-Saxon country clusters than between the ESEA country cluster and either of the other two. Hence, it was often rather difficult to explain differences between seemingly similar analytical entities or explain differences between them. It would have been desirable to include variables measured at the cluster level into the statistical analysis, which was not possible for mainly two reasons: Firstly, some of the information like characteristics of the school systems were not available as quantitative data. Secondly, due to the small number of level three entities (three clusters) the appropriate analytical method would have been Bayesian statistics, as traditional multilevel analysis yields unreliable results in the case of a small number of higher-

level units (Hox et al., 2012). Bayesian statistics, however, require larger computational power than was available for the large data set (about 29.000 teachers in about 1.600 schools were included in the 3 country clusters). Hence the decision to derive the tentative conclusions based on differences between the country clusters from the theoretical framework on country values and characteristics of the school systems and not include predictors on the country level. This, nonetheless, would be a fruitful direction for future research.

Furthermore, the two western clusters (Nordic and Anglo-Saxon countries) revealed more similarities with each other than with the ESEA countries. Accordingly, there seems to be a sub-divide within the sample: the two western clusters are culturally and normatively probably closer to each other than the eastern cluster, which is likely to be influenced by the historical legacy of Confucianism, which is not the case for the other two clusters. Between the Anglo-Saxon and Nordic cluster, the differences seem to be less notable and more difficult to explain, as the differences were developed in often comparatively similar cultural settings with features like the development of their political systems or welfare states. This might explain why, for some occasions, similar tentative explanations could be suggested for different types of TSE or countries. There were, for instance, more similarities in relation to classroom predictors between the Anglo-Saxon and Nordic countries, which could be explained by a similar degree of performance orientation supported in the countries, or school related features that might be perceived similarly due to a similar degree of power distance.

Another advantage of the large sample size offered by the OECD data-set is number of predictors it can accommodate. While many studies suffer from a too little sample in relation to the predictors that are of interest, it was possible to use a set of 23 predictors in both research questions. Tabachnick & Fidell (2013) suggest using 50 + 8*m (with m = number of predictors) as a rule of thumb to estimate the minimum sample size on each level for the desired number of predictors. Thus, even for the smallest analytical entity, the Anglo-Saxon cluster with $n_{teachers}$ = 6,477 and $n_{schools}$ = 397, allows the implementation of 15 predictors on the teacher- and classroom level and 8 predictors on the principal- and school level[16].

[16] Teacher predictors: 50+8*15 = 170

8.2.2 MISSING DATA TREATMENT

As discussed in Sections 3.8 and 3.9, missing data in large-scale studies is often inevitable, but there are modern methods to try to model this. In questionnaire data such as that used in this study, respondent fatigue, for example, can lead to attrition and limits the number of complete cases for analysis. In the complete TALIS data set, which is the result of merging a questionnaire filled in by teachers and a questionnaire filled in by the respective school principals, imputation was implemented in multiple steps. The merging was possible without any further loss of data, as there were individual identification numbers that allowed linking schools and teachers.

The missing data itself were thoroughly analysed with respect to their patterns and mechanisms using the appropriate tools in SPSS (see Section 3.10). As a result, multilevel multiple imputation was utilised as an appropriate strategy to account for missingness, as the estimates derived are based on the available observations for each participant (Schafer& Graham, 2002). This leads to more precise estimates and increased robustness even in the case of non-ignorable missing data. At the same time, multiple imputation is considered the state-of-the art in missing data treatment to date.

Another challenge was that the hierarchical structure of the data that theoretically required a three-level imputation model (teachers nested in schools nested in countries), which was not available in MPlus. This was solved by using an unrestricted two-level imputation model instead of a three-level imputation model and the data was imputed for each country separately and merged into one big data set afterwards. Using this imputation method drew on a thoroughly developed and reviewed method of accounting for missingness in large-scale data sets and contributes to yielding robust results (c.f. section 3.9). Furthermore, a relative efficiency of the imputed data of 96.32% was reached, which is deemed sufficient and allows for comparatively accurate analyses (Schafer & Graham, 2002).

Principal predictors: 50+8*8 = 114

8.2.3 A MULTI-FACETED MEASURE

Thirdly, in relation to this, although the three-dimensional measure of TSE taking into account the teacher's sense of self-efficacy in student engagement, instruction and classroom management is well established, there is still some debate whether TSE is a universal teacher trait that 'you either have or not have' or if it is a multidimensional trait that can differ within teachers (see Section 2.2). The differential relationships that the predictors have with the three dimensions on the one hand and the results of the measurement invariance testing on the other hand seem to contradict this idea for the present study and provide arguments for a more refined model of TSE, and the existence of more than the three domains. It was possible in this study to find support for the three domains even internationally; both across the 32 OECD countries and also within each of the three country clusters comprising Anglo-Saxon, Nordic and East and South-East Asian countries. My study's results thus support the multidimensional view of TSE. Furthermore, the tripartite instrument has not previously been used in a comparative analysis, which is why the current thesis contributes to the international development of the field with both new theoretical and empirical findings.

8.2.4 A STATE-OF-THE-ART INSTRUMENT

Over the past 40 years of research on TSE, a myriad of instruments has been used that conceptualised teacher's self-efficacy in different ways. As Klassen et al. (2010) pointed out in their review, some of those instruments suffer from a lack of conceptual clarity or do not strictly align with Bandura's definition of TSE. However, they are, nonetheless, part of the body of research on TSE. But it should be mentioned that despite these problematic measures used in TSE research, there are also some measures that show greater congruence with Bandura's work. The measure that is most closely linked and at the same time most widely used (as compared to Bandura-based, but rarely used) construct is the tripartite measure by Tschannen-Moran & Woolfolk-Hoy (2001), which was also used in this study. Hence, one of the difficulties emerging from this with respect to the present study was to disentangle the different ways in which the three dimensions: teachers' self-efficacy in instruction, student engagement and classroom management are related to previous research, individually. For instance, a future study using all three TSEs in a multivariate approach could answer whether

the predictors differentially predict different TSE domains and a profile analysis could be carried out. Sometimes, explanations turned out to be rather similar as for instance for the relationship between female principals and a higher sense of all types of TSE in the ESEA countries or the negative associations that private schools have with all types of TSE across all countries. Sometimes, in turn, it was possible to find more refined explanations that differed across the three types of TSE, as was the case, for instance, for the teachers' educational level and the lack of pedagogical training that teachers reported to be a problem in teacher education (König et al. 2011). This might either be because some features are more universally related to TSE as a whole (i.e. features of a private school), whereas other features are only related to certain aspects (i.e. community size, which was significantly related to ETSE in the Nordic countries, but not to MTSE. Future research should use the tripartite structure of TSE to gain deeper insight into how the different domains of TSE are related to features of the teachers, classrooms, principals and the school environment and contribute to further building a systematic corpus of research. Thus, it can be said that this study was able to contribute to further consolidate the field of research on teacher's self-efficacy because it used and further tested a measure comprising three domains that is currently accepted in the field and thoroughly based on theory.

8.2.5 A MULTILEVEL PERSPECTIVE

One specific characteristic of this data-set is that it pays respect to the hierarchical structure of the school environment. From a theoretical perspective this is a major benefit, as it allows the current study to take into account the perspective from the principal and the school environment, too, and thus contributes new knowledge to the field of research on TSE. Such multilevel analyses are still rare in TSE research (Klassen et al., 2010). This study was innovative in applying complex hierarchical models that require specific software and computational power to address the two research questions. The results show how TSE in three domains is predicted by teacher, classroom, principal and school characteristics and how the findings compare across domains and across the three country clusters. In addition to addressing the two main research questions, multilevel analysis was required to account for measurement invariance of the latent variables used to identify the three domains of TSE in instruction, student engagement and classroom management. Both, the use of multilevel techniques to

analyse the quality of the data and to test the hypothesis pay respect to the structure of the environment in which teachers work and thus address the phenomena studied more in-depth than previous studies have been able to. A limitation that is part of the use of questionnaire data is a bias in the subjective perspective that comes along with self-reported data, where social desirability often plays a role. Especially regarding classroom characteristics, the inclusion of outcome measures of student achievement or input measures of student background characteristics would have yielded additional and potentially more robust findings if achievement data or parents' reports or official school statistics could have been used instead of only the teacher's estimation of these student features in their schools and classes.

The multilevel models also helped disentangle variation across levels. Across both research questions and the three domains of TSE, it became apparent that most variation lies between teachers, least between schools. The first research question, which also investigated the variation across the 32 OECD countries taking part in TALIS, revealed that between 22% (ITSE) - 32% (ETSE) of the variation in TSE lies between countries. This finding was picked up in the second research question, where the three country clusters of Anglo-Saxon, Nordic and East- and South-East countries were compared. Here the focus was on how those predictors relate to TSE in the three domains in these three country clusters separately. This has contributed new findings to address the current gap in TSE research in three ways. Firstly, a representative international data-set was used. Secondly, the perspectives were advanced by modelling the principal and school measures as predictors. Thirdly, the multilevel perspective sheds light on how variation is distributed among teachers, schools and countries and plausible tentative explanations were proposed to account for variation of different levels.

8.2.6 BRINGING TOGETHER TEACHERS' AND PRINCIPAL'S PERSPECTIVES

This section somewhat ties in to the previous section about 'taking into account a multilevel perspective'. It differs from it in that sense that it does not concern the methodological aspects but consider the work environment of teachers. The majority of past studies in the literature have considered only teacher- and classroom predictors, whereas the school environment and principal characteristics have received little attention in TSE research so far (see literature review in Section 2.3.5 ff). To address this gap, information from the TALIS teacher

and principal questionnaires were utilised. As the teachers were reporting on their own biography, attitudes towards teaching and learning and the students and classrooms they are working in, and the principals on themselves, their schools and the wider school environment (OECD, 2013), a unique opportunity was created to investigate relationships between TSE across the three domains and the teachers' work environment. This contributes to knowledge on how TSE is related also to the wider work environment. In the second research question, it became apparent how different the associations are between different cultural settings, especially with regard to the principal- and school related predictors, which differed largely across the country clusters. This difference was particularly notable between the ESEA and the Nordic countries, where almost all versus almost none of the principal characteristics contributed to explaining variation in all three domains of teacher's self-efficacy in instruction, student engagement and classroom management.

8.2.7 BROADENING THE RESEARCH FOCUS

For a few cases the possible tentative explanations suggested for the association for the three dimensions of TSE and a particular predictor were rather similar (see i.e. principal leadership style). This might reflect two plausible reasons: the first is that the basis for intercultural research on TSE is rather thin. Thus, inferences were made across countries and cultural settings, as no direct information was available for that particular cultural context in the OECD TALIS data set. Future research could include country-level information from, for example, the GLOBE study (House et al., 2004), socio-demographic information or economic indicators like the Gini-coefficient or the GDP of a country. As has already been mentioned in the introduction to the thesis, the majority of studies on TSE have been conducted in the Anglo-Saxon countries and hardly any in the other countries in this study. This is especially true for the ESEA countries that as such have only been in the focus of research in TSE a few times (Malinen, 2016) and have many studies been published in Chinese. The Nordic countries, however, also lack a larger degree of country-specific research on TSE but are deemed to be culturally closer to the Anglo-Saxon countries, which makes comparisons easier. (There are studies by Skaalvik 2007, 2009, 2010, 2017, for example, that investigate TSE in Norway.) With this research, it was possible to start closing this gap by (1) providing new results that concern 32 OECD countries and (2) comparing and contrasting predictors of TSE in instruction, student

engagement and classroom management in three different country clusters. Through the second research question, it was possible to gain new insights into culture and school system specific characteristics of TSE, which in that manner have not yet been researched. Nonetheless, it is acknowledged that the interpretation of the results would have benefitted from a more comprehensive base of previous research on TSE in other countries.

8.3 DIRECTIONS FOR FUTURE RESEARCH

This study has been able to shed light on differences in teacher's perception of their own sense of self-efficacy in three domains (instruction, student engagement and classroom management). It has explored the country level (32 countries) and also three country clusters with respect to cultural differences and characteristics of the school systems. Given the instruments used and the insight gained from this research, four main possible avenues for further research emerge.

8.3.1 EXPANSION INTO OTHER COUNTRIES

For the expansion of research on TSE as a concept comprising three domains into other countries, two foci will be suggested. The first is to strengthen the cultural focus in research on TSE and to expand it to a large number of countries. Although this study has focused on 32 countries in the first and then 'zoomed into' three country clusters in the second research question, this can only be seen as a starting point. There is an apparent lack of cultural interpretations with regards to especially the wider school environment. Investigating this in the future could shed more light on differences identified in my research between the two western and one eastern country clusters that were not possible to explain with the evidence available on cultural values. The second is to employ the tri-partite measure of TSE more widely. For most past studies, only one-dimensional constructs of TSE have been used and only in recent years, the tripartite measure by Tschannen-Moran & Hoy (2001) has been used more often. Thus, explanations often required to make inferences from a rather general measure of TSE to a specific domain, as for example from general TSE to TSE in student engagement, whereas the newer studies allowed making more precise inferences. This would help in the future to disentangle the relationship the differences predictors have with the

three dimensions of TSE considered in my research and pay more attention to teaching as a multi-faceted task.

8.3.2 USING LINKS WITH OTHER CLASSROOM OBSERVATION TOOLS

One particular strength of the present tripartite construct is that it ties in with other well-established empirical measures of classroom characteristics as, for instance, the Classroom Assessment Scoring System (CLASS) observation tool, which aims at measuring and improving teacher effectiveness (Pianta, La Paro & Hamre, 2008) or the International System for Teacher Observation and Feedback (ISTOF) observation tool (Teddlie et al., 2006). Both of these seek to measure various aspects of teachers' and students' classroom behaviours and practices.

While the construct of TSE in its three domains focuses on teachers, CLASS is a systematic tool designed to observe the quality of student-teacher interactions with respect to the three domains of emotional support, instructional support and classroom management strategies of the teachers (Pianta, La Paro & Hamre, 2008). Within these three domains, CLASS measures twelve dimensions: a positive and negative climate, teacher sensitivity and regard for student perspective as part of emotional support, behaviour management, productivity and instructional learning formats as part of classroom organisation and finally content understanding, analysis and problem solving, quality of feedback and instructional dialogue as part of the instructional support the teacher provides. The same dimensions were used to define TSE in this study (c.f. Section 2.2.2). Hence, the results of my study could provide links with future research into the learning environment, because the measures are conceptually similar as they focus on instruction, classroom management and engagement of students in the lessons.

Another potential tool to link the TALIS TSE measure with is the International System for Teacher Observation and Feedback (ISTOF) (Teddlie et al., 2006). ISTOF measures five domains of teacher traits that contain eleven components. Those are Adaptive teaching (assessment and evaluation, differentiation and inclusion), quality of teaching (clarity of instruction, instructional skills, promoting active learning and developing metacognitive skills, planning of single lessons), long-term planning, classroom environment (classroom climate, classroom management) and teacher as a professional (teacher knowledge in subject, pedagogy and pedagogical content knowledge; teacher professionalism and reflectivity). The strengths of

ISTOF is that is an international framework to assess teacher effectiveness and provide feedback, regardless of national education systems. Thus, it is part of a strong international research base in the area of teacher effectiveness. As can be seen from the overlap between the three domains of TSE that TALIS covers and the domains and components of instruments such as CLASS and ISTOF, TALIS and its focus on the teacher and classroom from the teachers' perspective and the principal and school environment from the principals' perspective, which in the current data set covers 32 countries, could provide a fruitful addition to that by broadening the sources of information.

8.3.3 REACHING OUT TO NEW METHODS

Finally, the field of research on TSE could benefit from the use of methods other than quantitative ones: new studies that collect information from teachers and principals about cultural values and different TSE constructs using mixed methods approaches could investigate the topic in more detail than has been possible in this study that has relied on secondary data analysis and the limitations of the TALIS data set (see Sections 3.2 and 3.3). In the literature review it became apparent that most studies used some kind of quantitative methods (i.e. questionnaires, observational ratings, achievement data). This study also falls into the quantitative tradition of conducting research on TSE, which has mainly been followed so far, and there is a particular lack of longitudinal, qualitative or mixed methods studies. This has already been found by Klassen et al. (2010) in their meta-analysis and can also be concluded from the present study. Given the nature of claims that can be derived from quantitative research (Crotty, 1998), which offer an understanding of relationships between the domains of TSE and their characteristics, qualitative or mixed methods, in contrast, have the potential to offer insight into the reasoning for certain prevalence of TSE in countries or help to gain understanding in the development of it, which is still largely a black box. Especially with respect to differential results found for different domains of TSE within country clusters, as for example their relationship with the teachers' educational background or gender across clusters, the use of qualitative interviews, for example, could investigate further these differences in the teachers' perceptions for different domains of TSE.

Furthermore, this study is cross-sectional measuring one moment in time of a teacher's perception of their self-efficacy and thus no causal claims can be made but only associations can

be measured. Hence, future longitudinal studies that follow the development of TSE over time would be useful; as also noted by Zee & Koomen (2016) in their review.

8.3.4 COLLECTING INDEPENDENT DATA

As has already been mentioned in Sections 3.2 and 3.3, the use of OECD data and secondary data is not without concern. This is partly because of the economic scope within which the OECD collects its data, partly because of several limitations that come along with secondary data in general. When using such data-sets, the researcher is, for instance, beholden to the information for the quality of the data and has no control over the steps undertaken to choose instruments and monitor the data collection process. Thus, they are dependent on the reports provided by the OECD in this case and the research questions can only be specified with respect to the availability of subjects in the questionnaires. However, it, nonetheless, enabled me to make some generalisable claims across the 32 OECD countries due to the representative sample provided. For more individual research to address other gaps discovered in the literature, new data would need to be collected, for example, to use original student achievement data rather than teacher ratings. However, this data would then probably be of a much smaller scale, as the OECD is able to utilize large financial resources for breadth of coverage. For this study, the OECD provided a suitable data-set to analyse relevant research questions and to advance the geographical scope of the field of teacher's self-efficacy research.

As this data is part of the investigation into education systems initiated by the OECD, which would like to build a large data base offering a multi-perspective view, for some countries there are possible links between PISA test scores as a means of student achievement and TALIS data on the country level (OECD, 2013). There are, however, some concerns about using this approach. Firstly, the aggregation of teacher-, classroom-, principal- and school characteristics on the country level might only draw a rather blurry picture, as information on variation on the lower levels is lost. Secondly, there is more and more concern and critique raised regarding the quality and use of PISA data in education policy contexts, which is the ultimate aim of the OECD (OECD, 2013). One point of critique is the so called 'one-size-fits-all'-approach the OECD uses in terms of the PISA items that disregards the individual characteristics of the participating countries curricula. Thus, PISA data can only provide a cross-sectional

snapshot of the measured abilities of the tested 15-year-old students rather than an accurate description of the capability of the education systems (Baird et al., 2011). Although there are statistical methods to account for measurement related issues, the problems concerning the content of the items remain. Another issue raised is the selective use of the PISA results as a means to justify national decisions in education policy, which are often interpreted selectively and subjectively (Baird et al., 2016). Hence, despite some obvious advantages such as similarities in the approach of the education systems by the TALIS and the PISA study as part of OECD enquiry, the collection and use of independent assessment data that do greater justice to the respective education systems would be desirable. Furthermore, this would allow data-linkage on the teacher/classroom rather than country level and seems to be the more fruitful approach. It should be mentioned here that although the TALIS study is also part of international OECD enquiry, it has different limitations and is less affected by the PISA criticisms mentioned above since it has a less high stakes focus in accountability terms. In essence, TALIS can be seen as a large-scale study that uses psychometric scales or observable traits and thus is different in nature from the content-based attainment approach of the PISA surveys. See Section 3.3 for a more detailed discussion of the TALIS study in relation to the OECD.

8.4 IMPLICATIONS FOR EDUCATION POLICY

As with other empirical enquiries using OECD data, the findings of my study can be used to inform education policy and contribute to the advancement of education systems. There are, however, two challenges in identifying implications that stem from the international nature of the data-set. First, specific characteristics and needs from single education systems have not been considered. Second, the responsibilities for decisions concerning teacher training, classroom set-up or school policies might be held by different bodies in the different countries. Thus, the implications will be structured according to the field into which they fall, such as teacher training and education policy.

8.4.1 POSSIBLE IMPLICATIONS FOR TEACHER TRAINING

Teacher training can be seen as the initial phase and place of first contact with classroom teaching. The levers that might be relevant to inform teacher training are mainly found among the teacher- and classroom characteristics analysed.

General implications will be suggested for those aspects that had an omnibus association with TSE. Individual recommendations will be made for those instances, where differential results were found across domains or countries. The differential results for the country clusters will be suggested as a means to learn from successful aspects of the teacher's work environment in different settings.

1. Across all three domains of TSE, instruction, student engagement and classroom management, teaching experience showed a positive association. Although teaching experience is mainly the result of consecutive years of work experience, it would be helpful to provide student teachers with the opportunity to gain positive first classroom experiences. This suggestion is supported by Pinnock (2006) and Tucker et al., (2010) who found that opportunities for professional development are related not only to a higher level of TSE, but also by TSE theory (Bandura, 1977; 1996): professional development or supervision offers a source of vicarious experience as teachers can learn from a more experienced colleague or expert, a source of social persuasion as it offers the opportunity for constructive feedback and mastery experience, as the teachers awareness for classroom situations can be sensitised. This has the benefit that with the help of appropriate supervision and feedback, novice teachers may develop their self-efficacy across all three domains in a still relatively safe environment, before they have to face classroom situations on their own. The same implication can be derived from the second research question, where the three country clusters were compared, as the result was the same across all three domains of TSE and all three clusters.

2. In a similar vein, constructivist beliefs about teaching that take on a shared responsibility for student learning, seem to be relevant to contributing to a higher level of TSE internationally and across the three very different country clusters. Hence, a focus in teacher training on student-centred teaching methods that foster students' learning through their own inquiry and critical thinking would seem desirable to support teacher's TSE development.

3. Incorporating a gender perspective in teacher training could be suggested by the finding that female teachers in all analyses and domains (except MTSE in the ESEA

countries) reported a higher level of TSE. It seems to be worthwhile to aim at closing the gender gap here and make sure all students get to benefit from a high-TSE teacher.

4. With respect to the teacher's educational level, different findings came up for the three domains of TSE internationally and across the clusters. Internationally, it was positively related to TSE in instruction, negatively related to TSE in classroom management, but not significantly related to TSE in student engagement. Similarly, it was not related to TSE in student engagement across all country clusters, but positively to TSE in instruction in the ESEA countries and negatively to TSE in classroom management in the Anglo-Saxon clusters. As the answer cannot be by any means to lower educational standards for teachers, as the ultimate goal is highly educated students, teaching teachers instructional strategies, strategies on how to engage students in the lessons and how to manage a classroom successfully seems to lead teacher training in the right direction. This is in accordance with research by König et al. (2011) who found that pedagogy in teacher training is still deemed too little and teachers tend to feel not very well prepared for classroom tasks. As teachers might realise certain shortcomings of their teacher training or develop the need for special skills or knowledge only over the course of their career, in addition to strengthen the pedagogic modules in initial teacher training, offering flexible professional development opportunities seem to be worthwhile (see also point 1).

8.4.2 SUGGESTIONS FOR EDUCATION POLICY

Results from the present study reveal that the principal and the school environment also contribute to predicting teacher's self-efficacy. With respect to the three domains of TSE investigated and the cultural context, differences were observed that give hints for education policy makers on how to use different levers to foster teachers' sense of TSE across different domains.

1. Overall, the principal's instructional leadership style appeared to be most important in terms of principal characteristics to contributing to a high level of TSE. There were some differences with respect to country cluster and domain of TSE (i.e. there was no relationship between ITSE and the Anglo-Saxon cluster MTSE and ETSE in the

Nordic and Anglo-Saxon cluster). However, strong instructional leadership seems to be a desirable trait for principals to have in order to promote an environment in which TSE can flourish. This has also been shown by previous studies (Kirk, 2016; Fackler & Malmberg, 2016; Guo et al., 2011; Chester & Beaudin, 1996; Raudenbush et al., 1992). For education policy, this could mean making sure that principals receive appropriate training and development opportunities. An instructional style of leadership is associated with academic goal setting and a clear vision for the school (OECD, 2013), which can on the one hand, put pressure on the principal and on the other hand, requires a certain skillset. A focus on the development and monitoring of these skills seems to be a worthwhile effort so that teachers experience principals that shape the school environment with appropriate leadership skills.

2. In terms of the school characteristics, community size and the status as a public or private school, showed a significant relationship with TSE. Especially with respect to the community size, which serves as an indicator for the wider social environment in which the school is located, there are possible levers for education policy and politics. As has been discussed, there are certain characteristics associated with a rural or urban environment, such as a highly educated work force or difficult social environments (Copus et al., 2006). These might place different demands on the teachers and schools, as the parental support or attitude towards school can differ across these social groups (see i.e. Strand et al., 2011). Thus, it might be helpful for education policy to cater for the schools (and teachers) respective needs to help create support structures that foster an environment for TSE to flourish.

3. With respect to the type of school, it may be more difficult for education politics to have a direct influence. Private schools, especially, seem to provide a rather negative environment for the development of a high level of TSE internationally and those schools may be out of the reach of direct education policy influence. The findings suggest that they may be less attractive places to develop a positive sense of TSE in the three domains measured in this research.

8.5 SUMMARY

As has been discussed in this chapter, the results of the present study provide a number of pointers for future studies in teacher's self-efficacy. It encourages the use of TSE as a multi-faceted construct and the investigation into national school systems. It reveals consistent predictors across 32 countries and also produces findings on three different country clusters. The secondary analysis of TALIS data makes a contribution to new knowledge on theoretical and empirical findings and is the first large scale investigation to model country, school and teacher levels in exploring the three domains of TSE. The strengths and limitations of the study have been noted, as well as possible directions for further research. Some suggested implications of the research findings for education policy have also been proposed. There is, on the one hand, the opportunity to inform education policy and politics nationally by focussing on country results, but also to learn from successful mechanisms in other countries.

REFERENCES

Abumiya, M. (2012). *Upper Secondary Education in Japan*. Retrieved from https://www.nier.go.jp/English/educationjapan/pdf/201209SE.pdf

Aldridge, J. M., Fraser, B. J., & Huang, T. C. I. (1999). Investigating classroom environments in Taiwan and Australia with multiple research methods. *Journal of Educational Research*, *93*(1), 48–62. http://doi.org/10.1080/00220679909597628

Alivernini, F., & Lucidi, F. (2011). Relationship Between Social Context, Self-Efficacy, Motivation, Academic Achievement, and Intention to Drop Out of High School: A Longitudinal Study. *The Journal of Educational Research*, *104*(February), 241–252. http://doi.org/10.1080/00220671003728062

Allison, P. (2012). Missing Data. In *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 72–91). SAGE Publications Ltd. http://doi.org/10.4135/9780857020994

Andridge, R. R., & Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, *78*(1), 40–64. http://doi.org/10.1111/j.1751-5823.2010.00103.x

Antikainen, A. (2006). In search of the Nordic model in education. *Scandinavian Journal of Educational Research*, *50*(3), 229–243. http://doi.org/10.1080/00313830600743258

Appleby, G., & Lim, B. (2018). *Australian Constitutional Values*. (R. Dixon, Ed.). Oxford: Hart Pubishing.

Armor, D., Conry-Oseguera, P., & Cox, M. (1976). Analysis of the school prererred reading program in selected Los Angeles minority schools. *REPORT NO. R-2007- LAUSD*, *Santa Moni*.

Ashford, J. B., LeCroy, C. W., & Lortie, K. L. (2009). *Human Behavior in the Social Environment: A Multidimensional Perspective: A Multidimensional Perspective*. Independence: Cengage Learning.

Ashton, P. T., Doda, N., & Webb, R. B. (1983). *A study of teachers' sense of efficacy*. Gainesville.

Aslan, B. (2015). A Comparative Study on the Teaching Profession in Turkey and South Korea: Secondary Analysis of TALIS 2008 Data in Relation to Teacher Self-Efficacy. *Eurasian Journal of Educational Research*, (61), 1–22. http://doi.org/10.14689/ejer.2015.61.1

Augoustinos, M., Walker, I., & Donaghue, N. (2006). *Social Cognition: An Integrated Introduction*. Pine Forge Press.

Baird, J.-A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. (2011). Policy Effects of PISA. *Oxford University Centre for Educational Assessment*, (May), 41.

Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, *58*(2), 121–138. http://doi.org/10.1080/00131881.2016.1165410

Bandura, A. (1971). *Social learning theory*. General Learning Corporation.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. (E. Cliffs, Ed.)*PrenticeHall series in social learning theory* (Vol. 1). New Jersey: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. *Personnel Psychology* (Vol. 50). New York: Freeman.

Baraldi, A. N., & Enders, C. K. (2013). Missing Data Methods. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 1–34). Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199934898.013.0027

Boulton-Lewis, G. M., Smith, D. J. H., McCrindle, A. R., Burnett, P. C., & Campbell, K. J. (2001). Secondary teachers' conceptions of teaching and learning. *Learning and Instruction*, *11*(1), 35–51. http://doi.org/10.1016/S0959-4752(00)00014-1

Brown, T. (2013). Latent Variable Measurement Models. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 257–266). Oxford: Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199934898.013.0013

Bryman, A. (2001). The nature of quantitative research, 61–81.

Bryman, A. (2012). *Social Research Methods*. Oxford University Press.

Bulle, N. (2011). Comparing OECD educational models through the prism of PISA. *Comparative Education*, *47*(4), 503–521. http://doi.org/10.1080/03050068.2011.555117

Cajkler, W., & Hall, B. (2009). ' When they first come in what do you do ?' English as an additional language and newly qualified teachers. *Language and Education*, *23*(2), 153–170. http://doi.org/10.1080/09500780802308851

Calik, T., Sezgin, F., Kavgaci, H., & Kilinc, A. C. (2012). Examination of Relationships between Instructional Leadership of School Principals and Self-Efficacy of Teachers and Collective Teacher Efficacy. *Educational Sciences: Theory & Practice*, *12*(4), 2498–2504.

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy Beliefs as Determinants of Teachers' Job Satisfaction. *Journal of Educational Psychology*, *95*(4), 821–832.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs

as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, *44*(6), 473–490. http://doi.org/10.1016/j.jsp.2006.09.001

Chan, D. W. (2008a). General, collective, and domain-specific teacher self-efficacy among Chinese prospective and in-service teachers in Hong Kong. *Teaching and Teacher Education*, *24*(4), 1057–1069.

Chan, D. W. (2008b). Teacher self-efficacy and successful intelligence among Chinese secondary school teachers in Hong Kong. *Educational Psychology*, *28*(7), 735–746. http://doi.org/10.1080/01443410802259246

Chen, F., Bollen, K. a, Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research*, *29*(4), 468–508. http://doi.org/10.1177/0049124101029004003

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review*, *15*, 1–16. http://doi.org/10.1016/j.edurev.2015.02.001

Chester, M. D., & Beaudin, B. Q. (1996). Efficacy Beliefs of Newly Hired Teachers in Urban Schools. *American Educational Research Journal*, *33*(1), 233–257.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. http://doi.org/10.1207/S15328007SEM0902

Coladarci, T. (1992). Teachers' Sense of Efficacy and Commitment to Teaching. *The Journal of Experimental Education*, *4*(40), 323–337.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. http://doi.org/10.1037//1082-989X.6.4.330

Copus, A., Hall, C., Barnes, A., Dalton, G., Cook, P., Weingarten, P., … Irpud, A. H. (2006). *Study on Employment in Rural Areas*.

Cousins, B., Ross, J., & Gadalla, T. (1996). Within-teacher predictors of teacher efficacy. *Teaching and Teacher Education*, *12*(4), 385–400.

Crotty, M. (1998). *The Foundations of Social Research*. *The foundations of scial reserach meanings and perspectives in the research process*. SAGE.

Dewey, J. (1899). *The school and society*. Chicage: University of Chicago Press.

Dewey, J. (1902). *The child and the curriculum*. Chicago: University of Chicago Press.

Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.

Eagly, A., Karau, S., & Johnson, B. (1992). Gender and Leadership Style among School Principals: A Meta-Analysis. *Eduational Administration Quarterly*, *28*(1), 76–102.

Eberle, W. M. (2011). *Teacher self-efficacy and student achievement as measured by North Carolina reading and math end-of-grade tests*. *Electronic Theses and Dissertations.* East Tennessee State University.

Emmer, E. T., & Hickman, J. (1991). Teacher Efficacy in Classroom Management and Discipline. *Educational and Psychological Measurement*, *51*(3), 755–765.

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–138. http://doi.org/10.1037/1082-989X.12.2.121

European Commission/EACEA/Eurydice. (2014). *The Teaching and Learning International Survey 2013: Main findings from the survey and implications for education and training policies in Europe*. European Commission.

Fackler, S., & Malmberg, L.-E. (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects. *Teaching and Teacher Education*, *56*. http://doi.org/10.1016/j.tate.2016.03.002

Galton, M., Hargreaves, L., & Pell, T. (2009). Group work and whole - class teaching with 11 - to 14 - year - olds compared. *Cambridge Journal of Education*, *39*(1), 119–140. http://doi.org/10.1080/03057640802701994

Gardner, D. (2014). *Confucianism: A Very Short Introduction*. Oxford: Oxford University Press. http://doi.org/10.1093/actrade/9780195398915.001.0001

Geerlings, J., Thijs, J., & Verkuyten, M. (2018). Teaching in ethnically diverse classrooms: Examining individual differences in teacher self-efficacy. *Journal of School Psychology*, *67*(August 2017), 134–147. http://doi.org/10.1016/j.jsp.2017.12.001

George, S. V, Richardson, P. W., & Watt, H. M. (2018). Early career teachers' self-efficacy: A longitudinal study from Australia. *Australian Journal of Education*, *62*(2), 217–233. http://doi.org/10.1177/0004944118779601

Giallo, R., & Little, E. (2003). Classroom behaviour problems: The relationship between preparedness, classroom experiences, and self-efficacy in graduate and student teachers. *Australian Journal of Educational & Developmental Psychology*, *3*, 21–34.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569–582.

Gkolia, A., Dimitrios, B. A., & Koustelios, A. (2016). Background characteristics as predictors of Greek teachers' self-efficacy. *International Journal of Educational Management*, *30*(3), 460–472. http://doi.org/10.1108/IJEM-03-2014-0040

Goddard, R., Goddard, Y., Sook Kim, E., & Miller, R. (2015). A Theoretical and Empirical Analysis of the Roles of Instructional Leadership, Teacher Collaboration, and Collective Efficacy Beliefs in Support of Student Learning. *American Journal of Education*, *121*(4), 501–530. http://doi.org/10.1086/681925

Goldstein, H. (1995). Hierarchical Data Modeling in the Social Sciences. *Journal of Educational and Behavioral Statistics*, *20*(2), 201–204. http://doi.org/10.2307/1165357

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–76. http://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007b). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213. http://doi.org/10.1007/s11121-007-0070-9

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007a). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213. http://doi.org/10.1007/s11121-007-0070-9

Guo, Y., Justice, L. M., Sawyer, B., & Tompkins, V. (2011). Exploring factors related to preschool teachers' self-efficacy. *Teaching and Teacher Education*, *27*(5), 961–968.

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, *26*(4), 1094–1103. Retrieved from http://dx.doi.org/10.1016/j.tate.2009.11.005

Guskey, T. (1987). Context Variables That Affect Measures of Teacher Efficacy. *The Journal of Educational Research*, *1*(81), 41–47.

Guskey, T. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, *4*(1), 63–69.

Guskey, T., & Passaro, P. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, *31*(3), 627–643.

Guskey, T. R. (1981). Measurement of the Responsibility Teachers Assume For Academic Successes and Failures in the Classroom. *Journal of Teacher Education*, *32*(3), 44–51.

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, *44*(4), 932–942. http://doi.org/10.1016/j.paid.2007.10.034

Hanushek, E. a., & Woessmann, L. (2007). *Education Quality and Economic Growth*. The International Bank for Reconstruction and Development. Retrieved from www.worldbank.org

Hargreaves, L. (2009). The Status and Prestige of Teachers and Teaching. In L. J. Saha & A. G. Dworkin (Eds.), *International Handbook of Research on Teachers and Teaching* (pp. 217–229). New York: Springer.

Heider, F. (1958). *The psychology of interpersonal relations*. (F. Heider, Ed.)*The psychology of interpersonal relations* (Vol. New York). Wiley.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918. http://doi.org/10.1037/0022-3514.82.6.903

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, *37*(3), 137–150.

Hipp, K. A. (1996). Teacher Efficacy: Influence of Principal Leadership Behavior. Paper Presented at the Annual Meeting of the New York City Department of Educational Leadership. New York.

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, *10*(3), 301–320. http://doi.org/10.1016/0147-1767(86)90015-5

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, *2*(1), 1–26. http://doi.org/10.9707/2307-0919.1014

Hofstede, G. H., & Hofstede, G. J. (2010). *Cultures and organizations: software of the mind* (3rd ed.). New York: McGraw-Hill Education.

Holzberger, D., Philipp, A., & Kunter, M. (2014). Predicting teachers' instructional behaviors: The interplay between self-efficacy and intrinsic needs. *Contemporary Educational Psychology*, *39*(2), 100–111.

Hopfenbeck, T. N., Lenkeit, J., Masri, Y. El, Cantrell, K., Ryan, J., Baird, J., … Baird, J. (2017). Lessons Learned from PISA : A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, (January), 1–21. http://doi.org/10.1080/00313831.2016.1258726

House, R., Hanges, P., Javidan, M., Dorman, P., & Gupta, V. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. SAGE Publications, Inc.

Hox, J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. (G. A. Marcoulides, Ed.)*Quantitative methodology series* (Vol. 98). Routledge.

Hox, J. J. (2013). Multilevel Regression and Multilevel Structural Equation Modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 1–16). Oxford. http://doi.org/10.1093/oxfordhb/9780199934898.013.0014

Hox, J., van de Schoot, R., Matthijsse, S., & others. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*(2), 87–93.

Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*(4), 343–356.

Hoy, W., & Woolfolk, A. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, *93*(4), 355–372.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of Consumer Psychology*, *19*(4), 673–680. http://doi.org/10.1016/j.jcps.2009.09.002

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98. http://doi.org/10.1016/j.jcps.2009.09.003

Jarvis, P. (2003). *The Theory and Practice of Learning* (Vol. 8). London: Routledge.

Jerusalem, M., & Schwarzer, R. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. http://doi.org/10.1016/j.chb.2013.07.051

Kirk, J. M. (2016). *Principal leadership and teachers ' sense of self -efficacy : A meta-analysis*. University of Virginia.

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, *34*(1), 67–76. http://doi.org/10.1016/j.cedpsych.2008.08.001

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741–756.

Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology*, *36*(2), 114–129. http://doi.org/10.1016/j.cedpsych.2011.01.002

Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2010). Teacher Efficacy Research 1998–2009: Signs of Progress or Unfulfilled Promise? *Educational Psychology Review*, *23*(1), 21–43.

Klette, K., Sahlström, F., Blikstad-balas, M., Luoto, J., Tengberg, M., Roe, A., … Slotte, A. (2018). Justice through participation : student engagement in Nordic classrooms. *Education Inquiry*, *9*(1), 57–77. http://doi.org/10.1080/20004508.2018.1428036

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. *Analysis* (Vol. 77). http://doi.org/10.1038/156278a0

Knoblauch, D., & Woolfolk Hoy, A. (2008). "Maybe I can teach those kids." The influence of contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education*, *24*(1), 166–179.

Ko, J., Sammons, P., & Bakkum, L. (2016). *Effective Teaching*. Oxford.

König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F.-J. (2011). General Pedagogical Knowledge of Future Middle School Teachers: On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, *62*(2), 188–201. http://doi.org/10.1177/0022487110388664

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*(30), 1–21. http://doi.org/10.1207/s15327906mbr3001_1

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the Prekindergarten Year. *The Elementary School Journal*, *104*(5), 409–426. http://doi.org/10.1086/499760

Larsen, R. (2011). Missing Data Imputation versus Full Information Maximum Likelihood with Second-Level Dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(November), 649–662.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing and Health*, *25*(4), 295–306. http://doi.org/10.1002/nur.10041

Lee, M. N. N. (1999). Education in Malaysia : Towards Vision 2020. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *10*(1), 86–98. http://doi.org/10.1076/sesi.10.1.86.3514

Leonard, D. C. (2002). *Learning Theories: A to Z*. Greenwood Publishing Group.

Lim, S. C. (2007). Characteristics of mathematics teaching in Shanghai, China: Through the lens of a Malaysian. *Mathematics Education Research Journal*, *19*(1), 77–89. http://doi.org/10.1007/BF03217450

Lindorff, A., & Sammons, P. (2018). Going beyond structured observations: looking at classroom practice through a mixed method lens. *Zdm*, *50*(3), 521–534. http://doi.org/10.1007/s11858-018-0915-7

Little, T. D. (2014). *Longitudinal Structural Equation Modeling*. New York: The Guildford Press.

Malmberg, L. E., Hagger, H., & Webster, S. (2014). Teachers' situation-specific mastery experiences: teacher, student group and lesson effects. *European Journal of Psychology of Education*, *29*(3), 1–23.

McClelleand, D. C., Rindlisbacher, A., & DeCharms, R. (1955). Religious and other Sources of parental Attitudes toward independence Training. In D. C. McClelleand (Ed.), *Studies in Motivation* (pp. 389–397). New York: Appleton-Century-Croft.

Meredith, W. (1993). Measurement Invariance, Factor-Analysis and Factorial Invariance. *Psychometrika*, *58*(4), 525–543. http://doi.org/Doi 10.1007/Bf02294825

Meredith, W., & Teresi, J. A. (2006). An Essay on Measurement and Factorial Invariance. *Medical Care*, *44*(11, Suppl. 3), 699–713.

Merrell, K. W. (2003). *Behavioral, Social, and Emotional Assessment of Children and Adolescents*. Routledge.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247–258.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. *Statistical Approaches to Measurement Invariance*. Routledge. http://doi.org/10.4324/9780203821961

Mizell, H., Network, T., Staff, N., Council, D., National, G., Hotel, R., … Fellow, S. (2008). Self - Efficacy and Professional Development.

Mojavezi, A., & Tamiz, M. P. (2012). The Impact of Teacher Self-efficacy on the Students' Motivation and Achievement. *Theory and Practice in Language Studies*, *2*(3), 483–491. http://doi.org/10.4304/tpls.2.3.483-491

Mokshein, S. E., Zainuddin, Z., & Yusof, R. M. (2011). *Secondary Education Regional Information Base: Country Profile Malaysia*. *UNESCO Asia and Pacific Regional Bureau for Education*. Retrieved from http://uis.unesco.org/sites/default/files/documents/secondary-education-regional-information-base-country-profile-for-malaysia-en.pdf

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2014). *Handbook of Missing Data Methodology* (Vol. 6). Taylor & Francis Group.

Møller, J., & Skedsmo, G. (2013). Modernising education: New Public Management reform in the Norwegian education system. *Journal of Educational Administration and History*, *45*(4), 336–353. http://doi.org/10.1080/00220620.2013.822353

Morris, D. B., Usher, E. L., & Chen, J. A. (2017). Reconceptualizing the Sources of Teaching Self-Efficacy: a Critical Review of Emerging Literature. *Educational Psychology Review*, *29*(4), 795–833. http://doi.org/10.1007/s10648-016-9378-y

Muijs, D., Reynolds, D., Sammons, P., Kyriakides, L., Creemers, B. P. M., & Teddlie, C. (2018). Assessing individual lessons using a generic teacher observation instrument: how useful is the International System for Teacher Observation and Feedback (ISTOF)? *Zdm*, *0*(0), 0. http://doi.org/10.1007/s11858-018-0921-9

Muthén, L. K., & Muthén, B. O. (2012). *MPlus user's guide* (6th ed.). Los Angeles.

Norris, N. D. (2004). *The Promise and Failure of Progressive Education*. Oxford: Scarecrow Education.

OECD. (2004). *Completing the Foundation for Lifelong Learning An OECD Survey of Upper Secondary Schools*. Paris: OECD Publishing. http://doi.org/DOI:http://dx.doi.org/10.1787/9789264103733-en

OECD. (2011). *OECD work on education*. Paris: OECD Publishing.

OECD. (2012). Programme for International Student Assessment (PISA). OECD Publishing.

OECD. (2013a). About the OECD. Retrieved 10 July 2013, from http://www.oecd.org/about/

OECD. (2013b). OECD Teaching and Learning International Survey (TALIS) TALIS 2008. Retrieved 10 July 2013, from http://www.oecd.org/education/school/oecdteachingandlearninginternationalsurveytalistalis2008.htm

OECD. (2013c). *TALIS 2013 Technical Report*. Paris: OECD Publishing.

Oftedal Telhaug, A., Asbjørn Mediås, O., & Aasen, P. (2006). The Nordic model in education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research*, *50*(3), 245–283. http://doi.org/10.1080/00313830600743274

Paccagnella, O. (2006). Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. *Evaluation Review*, *30*(1), 66–85. http://doi.org/10.1177/0193841X05275649

Pang, N. S.-K. (2011). Educational Governance and Management in Sinic Societies. In Y. Zhao, J. Lei, G. Li, M. F. He, K. Okano, N. Megahed, … H. Ramanathan (Eds.), *Handbook of Asian Education: A Cultural Perspective*. New York: Routledge.

Parkay, F. W., Greenwood, G., Olejnik, S., & Proller, N. (1988). A study of the relationships among teacher efficacy, locus of control, and stress. *Journal of Research Development in Education*, *21*(4), 13–22.

Pas, E. T., Bradshaw, C. P., & Hershfeldt, P. a. (2012). Teacher- and school-level predictors of teacher efficacy and burnout: identifying potential areas for support. *Journal of School Psychology*, *50*(1), 129–45.

Peugh, J. L., & Enders, C. K. (2009). Review of Educational Research Missing Data in Educational Research : A Review of Reporting Practices and Suggestions for Improvement, *74*(4), 525–556. http://doi.org/10.3102/00346543074004525

Pinnock, A. (2006). *An Examination of the Effects of Professional Development on Teacher Educators' Attitudes and Self-Efficacy Towards Classroom Computer Technology*. Central Connecticut State University.

Pring, R. (2004). *Philosophy of Educational Research*. *Philosophy of educational research*. Continuum International Publishing Group.

Raudenbush, S., Rowan, B., & Cheong, Y. (1992). Contextual Effects on the Self-perceived Efficacy of High School Teachers. *Sociology of Education*, *65*(2), 150–167.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. (J. De Leeuw, Ed.)*Advanced Quantitative Techniques in the Social Sciences Series* (Vol. 1). Sage Publications.

Ravitch, D. (2008). America as a Historical Case Study Education and Democracy : The United States of. *Teachers College Record*, *110*(13), 42–57.

Rimm-Kaufman, S., & Sawyer, B. (2004). Primary-Grade Teachers' Self-Efficacy Beliefs, Attitudes toward Teaching, and Discipline and Teaching Practice Priorities in Relation to the Responsive Classroom Approach. *The Elementary School Journal*, *104*(4), 321–341.

Romi, S., Lewis, R., & Katz, Y. J. (2009). Student responsibility and classroom discipline in Australia, China, and Israel. *Compare*, *39*(4), 439–453. http://doi.org/10.1080/03057920802315916

Rose, J., & Medway, F. (1981a). Measurement of teachers' beliefs in their control over student outcomes. *Journal of Educational Research*, *74*(3).

Rose, J., & Medway, F. (1981b). Teacher locus of control, teacher behavior, and student behavior as determinants of student achievement. *The Journal of Educational Research*, *6*(74), 375–381.

Rosker, J. S. (2016). Modern Confucianism and the Concept of "Asian Values". *Asian Studies*, *4*(1), 153–164. http://doi.org/10.4312/as.2016.4.1.153-164

Ross, J. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue Canadienne de l …*, *1*(1977), 51–65.

Ross, J. A., & Bruce, C. D. (1997). *Effects of Professional Development on Teacher Efficacy: Results of a Randomizes Field Trial*.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173–220). Academic Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*(1), 1–28.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley-Interscience.

Rubin, D. B., & Little, R. J. A. (2014). *Statistical Analysis with Missing Data* (2nd ed.). New Jersey: John Wiley & Sons.

Rutkowski, D., Rutkowski, L., Bélanger, J., Knoll, S., Weatherby, K., & Prusinski, E. (2013). *TALIS 2013 Conceptual Framework*. Paris: OECD Publishing.

Ryu, E. (2015). Multiple group analysis in multilevel structural equation model across level 1 groups. *Multivariate Behavioral Research*, *50*(3), 300–15. http://doi.org/10.1080/00273171.2014.1003769

Sato, N., & McLaughlin, M. W. (1998). Context Matters: Teaching in Japan and in the United States. *Education and Training in Japan: Volume II*, *73*(5), 150–164.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Scheerens, J., Witziers, B., & Steen, R. (2013). A Meta-analysis of School Effectiveness Studies. *Revista de Educación*, (361), 619–645. http://doi.org/10.4438/1988-592X-RE-2013-361-235

Scherer, R., Jansen, M., Nilsen, T., Areepattamannil, S., & Marsh, H. W. (2016). The quest for comparability: Studying the invariance of the teachers' sense of self-efficacy (TSES) measure across countries. *PLoS ONE*, *11*(3), 1–29. http://doi.org/10.1371/journal.pone.0150829

Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology*, *57*(SUPPL. 1), 152–171. http://doi.org/10.1111/j.1464-0597.2008.00359.x

Schwarzer, R., & Schmitz, G. (2002). Individuelle und kollektive Selbstwirksamkeitserwartung von Lehrern. *Zeitschrift Für Pädagogik*, *44*(0), 192–214. http://doi.org/10.7863/jum.1989.8.4.211

Sjøberg, S. (2015). PISA and global educational governance - A critique of the project, its uses and implications. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(1), 111–127. http://doi.org/10.12973/eurasia.2015.1310a

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3), 611–625.

Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, *26*(4), 1059–1069.

Skaalvik, E. M., & Skaalvik, S. (2014). Teacher Self-Efficacy and Perceived Autonomy: Relations with Teacher Engagement, Job Satisfaction, and Emotional Exhaustion. *Psychological Reports*, *114*(1), 68–77. http://doi.org/10.2466/14.02.PR0.114k14w0

Skaalvik, E. M., & Skaalvik, S. (2017). Motivated for teaching? Associations with school goal structure, teacher self-efficacy, job satisfaction and emotional exhaustion. *Teaching and Teacher Education*, *67*, 152–160. http://doi.org/10.1016/j.tate.2017.06.006

Smith, E. (2008). *Using secondary data in educational and social research*. Open University Press.

Smithers, A., & Robinson, P. (2010). *Choice and Selection in School Admissions: the experience of other countries*. Retrieved from https://www.suttontrust.com/wp-content/uploads/2010/11/1smithers-final-report-1.pdf

Snijders, T. A. B. (2005). Power and Sample Size in Multilevel Linear Models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science. Volume 3* (pp. 1570–1573). Chicester: Wiley.

Snijders, T., & Bosker, R. (2011). *Multilevel modeling: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage Publications Ltd.

Speybroeck, S., Kuppens, S., van Damme, J., van Petegem, P., Lamote, C., Boonen, T., & de Bilde, J. (2012). The role of teachers' expectations in the association between children's SES and performance in kindergarten: A moderated mediation analysis. *PLoS ONE*, *7*(4), 1–8. http://doi.org/10.1371/journal.pone.0034502

Stevenson, H. W., & Lee, S. (1995). The East Asian Version of Whole-Class Teaching. *Educational Policy*, *9*(2), 152–168. http://doi.org/10.1177/0895904895009002004

Strand, S. (2015). Ethnicity, deprivation and educational achievement at age 16 in England: trends over time, (June). http://doi.org/978-1-78105-499-4

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. *Using multivariate statistics 5th ed*. Allyn & Bacon/Pearson Education.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation*, *12*(6), 561–582. http://doi.org/10.1080/13803610600874067

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

Tucker, C. M., Porter, T., Reinke, W. M., Herman, K. C., Ivery, P. D., Mack, C. E., & Jackson, E. S. (2005). Promoting Teacher Efficacy for Working With Culturally Diverse Students. *Preventing School Failure: Alternative Education for Children and Youth*, *50*(1), 29–34.

United Nations. (2017). *Trends and drivers of international migration in Asia and the Pacific* (Vol. 2). Bangkok. Retrieved from http://www.unescap.org/sites/default/files/GCMPREP_1E.pdf

van Buuren, S. (2011). Multiple Imputation of Multilevel Data. *Handbook of Advanced Multilevel Analysis*, 173–196.

Vanderberg, R. J., & Lance, C. E. (2000). A review an synthesis of the measurement in variance literature : Sugestions practice, and recommendations for organizational research. *Organizational Research Methods.*, *3*(1), 4–70. http://doi.org/10.1177/109442810031002

Vartanian, T. P. (2011). Secondary Data Analysis. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 1–216). Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195388817.001.0001

Vasquez, A. (2008). *Teacher efficacy and student achievement in ninth and tenth grade reading : A multilevel analysis*. University of South Florida. Retrieved from http://scholarcommons.usf.edu/etd/544

Vieluf, S., Kunter, M., & Vijver, F. J. R. Van De. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education*, (35), 92–103.

Vijver, F. J. Van De, & He, J. (2014). Report on Social Desirability , Midpoint and Extreme Responding in TALIS 2013, (107).

Walker, J., & Slear, S. (2011). The Impact of Principal Leadership Behaviors on the Efficacy of New and Experienced Middle School Teachers. *NASSP Bulletin*, *95*(1), 46–64.

Webster, R. S. (2017). Education or quality of teaching? Implications for australian democracy. *Australian Journal of Teacher Education*, *42*(9), 59–76. http://doi.org/10.14221/ajte.2017v42n9.4

West, M. R. (2010). "Every Catholic Child in a Catholic School ": Competition, and Student Achievement across Countries. *The Economic Journal*, *120*(May 2008), 229–255. http://doi.org/10.1111/J.1468-0297.2010.02375.X.

Westbury, I., Hansén, S., Kansanen, P., & Björkvist, O. (2005). Teacher Education for Research-based Practice in Expanded Roles: Finland's experience. *Scandinavian Journal of Educational Research*, *49*(5), 475–485. http://doi.org/10.1080/00313830500267937

Wolters, C. a., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, *99*(1), 181–193.

Woolfolk, a . E. ., & Hoy, W. . K. . (1990). Prospective teachers ' sense of efficacy and beliefs about control . Journal of Educational Psychology , 82 , 81-91 . Originally based on the Teacher Efficacy Scale developed by S . Gibson & M . Dembo (1. *Journal of Educational Psychology*, *82*, 82–91.

Woolfolk, A., & Hoy, W. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, *82*(1), 81–91.

Woolfolk, A., Rosoff, B., & Hoy, W. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education*, *6*(2), 137–148.

Yucel, R. M., & Zaslavsky, A. M. (2001). Practical suggestions on rounding in multiple imputation. *Evaluation*, 4679–4683.

Yun, J. T., & Reardon, S. F. (2005). Private School Racial Enrollments and Segregation. In J. Scott (Ed.), *School Choice and Diversity* (pp. 42–58). New York: Teachers College Press.

Zee, M., & Koomen, H. M. Y. (2016). Teacher Self-Efficacy and Its Effects on Classroom Processes, Student Academic Adjustment, and Teacher Well-Being: A Synthesis of 40 Years of Research. *Review of Educational Research*, *86*(4), 981–1015. http://doi.org/10.3102/0034654315626801

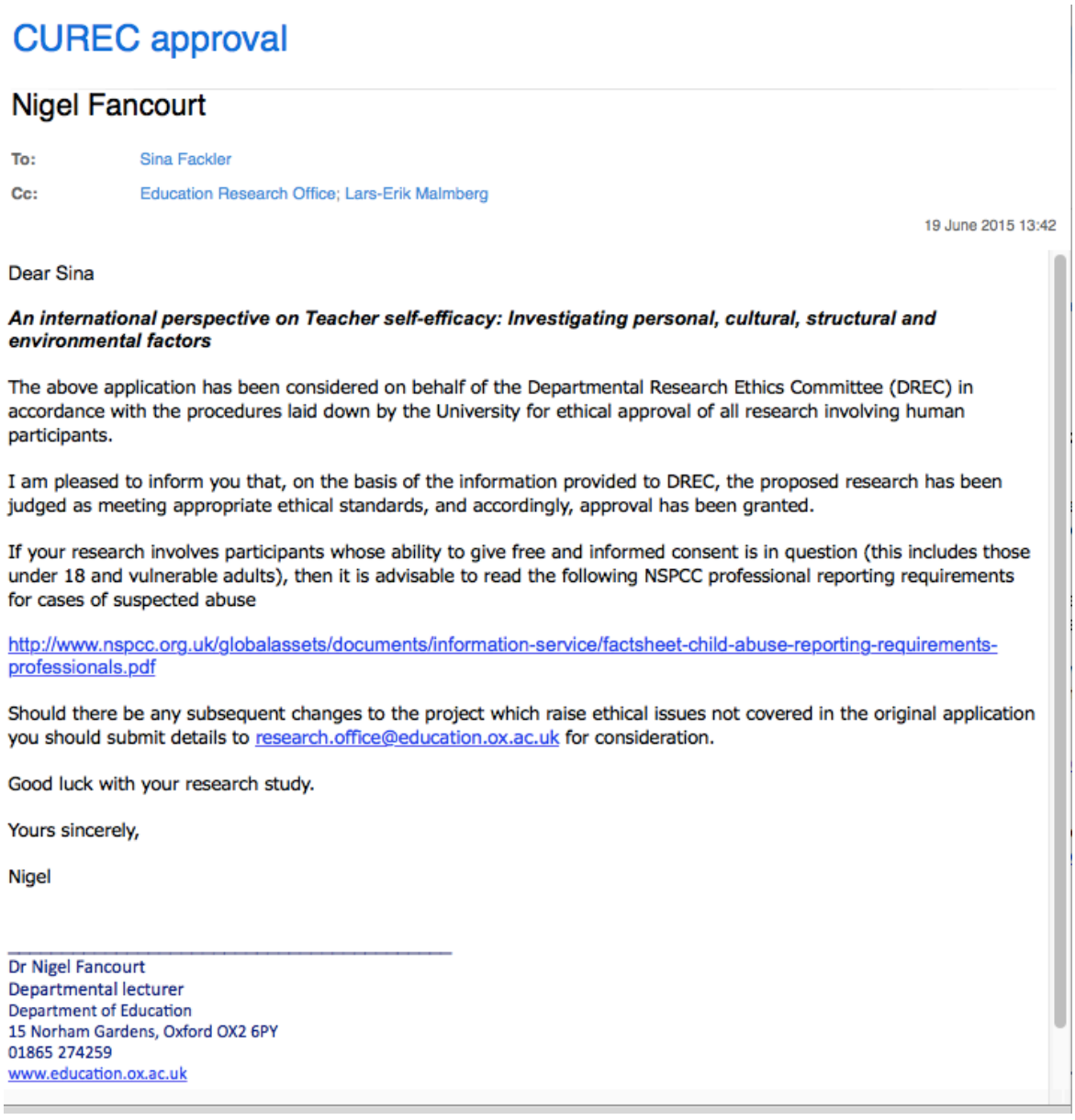
Zee, M., Koomen, H. M. Y., Jellesma, F. C., Geerlings, J., & de Jong, P. F. (2016). Inter- and intra-individual differences in teachers' self-efficacy: A multilevel factor exploration. *Journal of School Psychology*, *55*, 39–56. http://doi.org/10.1016/j.jsp.2015.12.003

Zhao, X., Selman, R. L., & Haste, H. (2015). Academic stress in Chinese schools and a proposed preventive intervention program. *Cogent Education*, *2*(1), 1–14. http://doi.org/10.1080/2331186X.2014.1000477

APPENDIX

The appendix contains a copy of the CUREC letter of approval for the study from the University of Oxford ethics committee, the descriptive statistics for the predictor variables of research questions 1 and 2 and normality statistics of the level 1 variables to show the data meet the criteria for multilevel regression. It furthermore contains a different type of analysis for the second research questions, which investigates whether the differences observed in the predictors of the three country clusters is significant.

Appendix 1: CUREC letter of approval



CUREC approval

Nigel Fancourt

To: Sina Fackler

Cc: Education Research Office; Lars-Erik Malmberg

19 June 2015 13:42

Dear Sina

An international perspective on Teacher self-efficacy: Investigating personal, cultural, structural and environmental factors

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse

http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Yours sincerely,

Nigel

Dr Nigel Fancourt
Departmental lecturer
Department of Education
15 Norham Gardens, Oxford OX2 6PY
01865 274259
www.education.ox.ac.uk

Appendix 2: Descriptive statistics RQ 1

This section contains a table with descriptive statistics for research question 1, comprising all 32 OECD countries in the sample. They are organised in the same way as the results tables, providing the descriptive properties of the data according to teacher-, classroom-, principal- and school characteristics analysed.

Table 30 Descriptive statistics for independent variables across 32 OECD countries

	N				SE			
Teacher	**Valid**	**Missing**	**%**	**Mean**	**Mean**	**SD**	**Skewness**	**Kurtosis**
Female Gender	104,285	3	.003					
Educational Level	102,841	1,447	1.407	2.907	.001	.393	-2.814	11.003
Teaching Exp.	97,709	6,579	6.733	16.143	.033	10.381	.508	-.594
Hours Teaching	101,433	2,855	2.815	20.016	.029	9.121	1.468	6.383
Hours Planning	100,249	4,039	4.029	7.103	.020	6.395	4.048	32.102
Hours Admin. Work	94,479	9,809	10.382	2.913	.015	4.636	7.326	92.114
Construct. Bel.	99,560	4,728	4.749	12.732	.006	1.935	-.198	.473
Classroom								
Class size	85,606	18,682	21.823	25.263	.034	9.855	1.081	5.274
Classroom Climate	85,989	18,299	21.281	11.312	.009	2.672	-.419	-.136
Low Student Achieve-ment	99,374	4,914	4.945	2.779	.003	.983	.565	-.200
Teaching & Learning	84,363	19,925	23.618	77.570	.057	16.503	-1.344	1.979
Keeping Order	84,363	19,925	23.618	13.288	.042	12.189	1.879	4.967
Admin. Tasks	84,363	19,925	23.618	8.429	.026	7.428	2.772	14.715
Different Language	99,483	4,805	4.830	1.779	.004	1.120	1.589	1.757
Disadv. Background	99,257	5,031	5.069	2.532	.004	1.110	.628	-.295
	104,288							
Principal								
Female Gender	6,090	330	5.419					
Age	6,071	349	5.749	50.609	.105	8.198	-.318	-.431
Educational Level	6,059	361	5.958	2.991	.004	.301	-1.280	14.878
Work Experience	5,704	716	12.553	8.684	.095	7.187	1.203	1.311
Instructional Leadership	6,005	415	6.911	8.632	.023	1.816	-.050	-.205
School								
Student Enrolment	5,993	427	7.125	661.078	6.375	493.555	1.474	3.438
Type of School	6,084	336	5.523					
Community Size	6,062	358	5.906	3.772	.018	1.424	-.211	-.709
	6,420							

Appendix 3: Descriptive statistics RQ 2

As for the predictors used in research question 1, the following three tables contain information on the descriptive properties. There are separate tables for each country cluster: Anglo-Saxon-, Nordic- and East & South-East Asian countries.

Table 31 Descriptive statistics for predictors in RQ2; Anglo-Saxon countries

Anglo-Saxon Countries

Teacher	Valid	N Missing	%	Mean	SE Mean	SD	Skewness	Kurtosis
Female Gender	6,476	1	.02					
Educational Level	6,443	34	.53	2.99	.003	.20	-4.79	60.81
Teaching Exp.	6,333	144	2.27	14.10	.128	10.16	.75	-.37
Hours Teaching	6,352	125	1.97	21.68	.118	9.43	1.29	5.68
Hours Planning	6,282	195	3.10	7.49	.075	5.92	3.10	19.75
Hours Admin. Work	6,165	312	5.06	3.93	.059	4.62	4.75	42.87
Construct. Bel.	6,092	385	6.32	12.65	.024	1.90	-.10	.26
Classroom								
Class size	5,182	1,295	24.99	25.42	.141	10.14	2.64	13.37
Classroom Climate	5,188	1,289	24.85	11.66	.038	2.75	-.49	-.10
Low Student Achievement	6,075	402	6.62	2.92	.014	1.11	.34	-.61
Teaching & Learning	5,166	1,311	25.38	79.54	.222	15.95	-1.46	2.33
Keeping Order	5,166	1,311	25.38	13.26	.184	13.24	1.97	4.73
Admin. Tasks	5,166	1,311	25.38	6.76	.086	6.19	3.46	25.20
Different Language	6,083	394	6.48	2.12	.015	1.15	1.20	.76
Disadv. Background	6,072	405	6.67	2.84	.015	1.18	.33	-.82
	6,477							
Principal								
Female Gender	366	31	8.47					
Age	363	34	9.37	50.54	.399	7.61	-.33	-.21
Educational Level	365	32	8.77	3.07	.015	.29	1.56	12.95
Work Experience	359	38	10.58	7.81	.307	5.82	1.10	1.36
Instructional Leadership	354	43	12.15	9.08	.094	1.76	.00	-.50
School								
Student Enrolment	356	41	11.52	973.88	24.687	465.80	.56	.39
Type of School	360	37	10.28					
Community Size	361	36	9.97	4.43	.067	1.28	-.52	-.18
	397							

Table 32 Descriptive statistics for predictors in RQ2; Nordic countries

Nordic Countries

Teacher	Valid	N Miss-ing	%	Mean	SE Mean	SD	Skew-ness	Kurto-sis
Female Gender	10,678	-	.00					
Educational Level	10,621	57	.54	2.93	.003	0.35	-4.09	19.03
Teaching Exp.	10,435	243	2.33	15.86	.105	10.71	0.64	-0.58
Hours Teaching	10,437	241	2.31	17.84	.059	6.00	-0.03	1.55
Hours Planning	10,287	391	3.80	6.31	.043	4.41	3.24	32.08
Hours Admin. Work	9,692	986	10.17	2.86	.034	3.38	5.16	66.64
Construct. Bel.	10,031	647	6.45	12.20	.018	1.79	-0.08	0.65
Classroom								
Class size	8,232	2,446	29.71	20.91	.100	9.09	2.68	16.85
Classroom Climate	8,327	2,351	28.23	11.41	.029	2.64	-0.40	-0.11
Low Student Achieve-ment	10,116	562	5.56	2.75	.009	0.93	0.81	0.35
Teaching & Learning	8,255	2,423	29.35	81.87	.148	13.47	-1.59	3.76
Keeping Order	8,255	2,423	29.35	10.98	.115	10.45	2.17	7.48
Admin. Tasks	8,255	2,423	29.35	6.70	.065	5.88	2.81	16.65
Different Language	10,141	537	5.30	2.00	.010	1.02	1.27	1.47
Disadv. Background	10,017	661	6.60	2.11	.009	0.86	0.95	1.41
	10,678							
Principal								
Female Gender	544	76	13.97					
Age	543	77	14.18	51.59	.326	7.60	-0.18	-0.85
Educational Level	544	76	13.97	2.97	.011	0.26	-3.41	26.64
Work Experience	534	86	16.10	9.67	.297	6.86	0.82	0.29
Instructional Leadership	539	81	15.03	7.58	.074	1.72	0.25	0.00
School								
Student Enrolment	541	79	14.60	374.88	8.826	205.29	1.01	2.05
Type of School	543	77	14.18					
Community Size	542	78	14.39	3.51	.054	1.25	-0.15	-0.58
	620							

Table 33 Descriptive statistics for predictors in RQ2; ESEA countires

East- & South-East Asian Countries

Teacher	Valid	N Miss-ing	%	Mean	SE Mean	SD	Skew-ness	Kurto-sis
Female Gender	12,510	-	.00					
Educational Level	12,480	30	.24	2.95	.003	0.29	-4.08	23.30
Teaching Exp.	12,267	243	1.98	14.40	.093	10.25	0.48	-0.78
Hours Teaching	12,419	91	.73	17.74	.070	7.79	3.24	21.64
Hours Planning	12,326	184	1.49	7.83	.064	7.16	3.42	20.28
Hours Admin. Work	12,209	301	2.47	5.56	.060	6.59	4.75	38.30
Construct. Bel.	12,286	224	1.82	12.86	.017	1.89	-0.03	0.33
Classroom								
Class size	10,058	2,452	24.38	33.17	.099	9.91	0.40	7.12
Classroom Climate	10,105	2,405	23.80	11.58	.026	2.57	-0.44	0.04
Low Student Achieve-ment	12,291	219	1.78	2.79	.009	0.99	0.67	-0.14
Teaching & Learning	9,911	2,599	26.22	74.36	.186	18.53	-1.03	0.72
Keeping Order	9,911	2,599	26.22	15.91	.131	13.08	1.58	3.27
Admin. Tasks	9,911	2,599	26.22	9.31	.094	9.33	2.18	7.46
Different Language	12,314	196	1.59	2.11	.012	1.36	0.93	-0.50
Disadv. Background	12,295	215	1.75	2.49	.009	0.99	0.65	0.04
	12,510							
Principal								
Female Gender	645	33	5.12					
Age	647	31	4.79	54.96	.225	5.72	-1.13	1.23
Educational Level	647	31	4.79	3.02	.007	0.17	1.42	49.66
Work Experience	644	34	5.28	5.46	.162	4.11	1.38	2.50
Instructional Leadership	637	41	6.44	8.73	.077	1.94	0.06	-0.84
School								
Student Enrolment	643	35	5.44	919.99	22.057	559.30	1.01	2.60
Type of School	645	33	5.12					
Community Size	645	33	5.12	4.80	.051	1.28	-0.91	-0.02
	678							

The following table (Table 5) shows the reliability statistics representing the internal consistency of the scales used for each country cluster. The scales yield high to very high internal consistency ranging from .71 (instructional leadership style/Nordic countries) to .93 (MTSE in the ESEA countries).

Table 34 Cronbach's alpha for each latent variable per country cluster

	α		
	Nordic	Anglo-Saxon	ESEA
ETSE	.80	.84	.89
ITSE	.79	.78	.76
MTSE	.84	.84	.93
Con. Bel.	.83	.93	.76
CR Climate	.92	.91	.93
Inst. Lead.	.71	.88	.86

Appendix 4: Correlation matrix of predictor variables

The following table shows the correlation matrix for each predictor with each predictor of the statistical models analysed in RQ1 and RQ2 to test, whether there is collinearity between variables used. The highest correlation coefficient found between any two variables is -.861** for the classroom time spent on teaching & learning and the classroom time spent on keeping order. Hence, strong multicollinearity for the variables in the models can be excluded (Tabatschnick & Fidell, 2007).

Table 35 Correlation matrix for predictor variables of RQ1 and RQ2

Correlation matrix

Pearson Correlation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1 Prin. Gender	1	-,080**	-,029**	-,024**	-.010	,063**	,062**	,091**	-,047**	-,024**	-,018**	-,016**	-,014*	,149**	,037**	,055**	-,037**	,029**	-.003	-.012	-,023**	-,063**	,141**
2 Prin. Age	-,080**	1	-,017**	,317**	,047**	,015*	-,026**	-,036**	,103**	,027**	-,054**	,021**	,055**	-,189**	-,077**	-,091**	,062**	-.013	-.007	.009	,020**	,064**	-,033**
3 Prin. Edulevel	-,029**	-,017**	1	.002	,034**	-,038**	,115**	.010	-.012	,061**	,056**	.008	-.012	-,035**	-.002	,035**	,027**	-.004	-.012	.013	,065**	.006	,138**
4 Prin. Workexp.	-,024**	,317**	.002	1	-,084**	-,044**	-.003	,012*	-.008	-.010	.008	-,031**	-,102**	,064**	,017**	-,045**	-,157**	-,033**	-,062**	,063**	-,013*	-.003	-,052**
5 School location	-.010	,047**	,034**	-,084**	1	-,062**	,342**	-.002	-,057**	,046**	-.007	,083**	,107**	,180**	.011	.007	,293**	,086**	,098**	-,114**	,087**	-,037**	,094**
6 School Type 'Private'	,063**	,015*	-,038**	-,044**	-,062**	1	-,026**	,047**	,014*	-,023**	.004	-.008	,026**	,079**	,101**	,137**	,077**	,060**	,080**	-,090**	-,019**	-,083**	,024**
7 School Enrol.	,062**	-,026**	,115**	-.003	,342**	-,026**	1	,063**	-,120**	,015*	,021**	,037**	,103**	,248**	,036**	,166**	,326**	,135**	,096**	-,134**	,089**	-,061**	,336**
8 Tea. Gender	,091**	-,036**	.010	,012*	-.002	,047**	,063**	1	-,055**	.005	.002	,014*	-.011	,038**	-.001	-.008	-,040**	-,015*	,015*	-.008	.003	-,041**	,061**
9 Tea. Workexp.	-,047**	,103**	-.012	-.008	-,057**	,014*	-,120**	-,055**	1	-,067**	-,055**	-,035**	,018**	-,103**	-,074**	-,105**	-,065**	-,122**	-,178**	,180**	-,080**	,146**	-,095**
10 Tea. Edulevel	-,024**	,027**	,061**	-.010	,046**	-,023**	,015*	.005	-,067**	1	,028**	,026**	,015*	-,062**	-.006	.000	,029**	-.005	.002	.003	,034**	.008	-.005
11 Tea. Hours Teaching	-,018**	-,054**	,056**	.008	-.007	.004	,021**	.002	-,055**	,028**	1	,106**	-,025**	-,013*	,063**	,116**	-.005	.007	,041**	-,034**	,038**	-,020**	,033**
12 Tea. Hours Planning	-,016**	,021**	.008	-,031**	,083**	-.008	,037**	,014*	-,035**	,026**	,106**	1	,285**	,012*	,020**	,012*	,058**	.011	,041**	-,037**	,042**	-.008	,015*
13 Tea. Hours Admin. Work	-,014*	,055**	-.012	-,102**	,107**	,026**	,103**	-.011	,018**	,015*	-,025**	,285**	1	,027**	,031**	,052**	,120**	,088**	,037**	-,073**	,017**	,015*	,066**
14 CR. Diff. Lang.	,149**	-,189**	-,035**	,064**	,180**	,079**	,248**	,038**	-,103**	-,062**	-,013*	,012*	,027**	1	,327**	,332**	,098**	,127**	,128**	-,151**	-.007	-,163**	,193**
15 CR. Low Achievers	,037**	-,077**	-.002	,017**	.011	,101**	,036**	-.001	-,074**	-.006	,063**	,020**	,031**	,327**	1	,523**	,031**	,116**	,305**	-,280**	.006	-,332**	,092**
16 CR. Disadv. Bckg.	,055**	,091**	,035**	-,045**	.007	,137**	,166**	-.008	-,105**	.000	,116**	,012*	,052**	,332**	,523**	1	,121**	,137**	,258**	-,257**	,054**	-,234**	,231**
17 CR. Class Size	-,037**	,062**	,027**	-,157**	,293**	,077**	,326**	-,040**	-,065**	,029**	-.005	,058**	,120**	,098**	,031**	,121**	1	,130**	,103**	-,135**	,077**	-,038**	,123**
18 CR. Admni Tasks.	,029**	-.013	-.004	-,033**	,086**	,060**	,135**	-,015*	-,122**	-.005	.007	.011	,088**	,127**	,116**	,137**	,130**	1	,268**	-,656**	.006	-,145**	,093**
19 CR. Keeping Order	-.003	-.007	-.012	-,062**	,098**	,080**	,096**	,015*	-,178**	.002	,041**	,041**	,037**	,128**	,305**	,258**	,103**	,268**	1	-,861**	,019**	-,531**	,059**
20 CR. Teaching Learning	-.012	.009	.013	,063**	-,114**	-,090**	-,134**	-.008	,180**	.003	-,034**	-,037**	-,073**	-,151**	-,280**	-,257**	-,135**	-,656**	-,861**	1	-.012	,473**	-,089**
21 Tea. Constr. Believes	-,023**	,020**	,065**	-,013*	,087**	-,019**	,089**	.003	-,080**	,034**	,038**	,042**	,017**	-.007	.006	,054**	,077**	.006	,019**	-.012	1	,016*	,070**
22 CR. Climate	-,063**	,064**	.006	-.003	-,037**	-,083**	-,061**	-,041**	,146**	.008	-,020**	-.008	,015*	-,163**	-,332**	-,234**	-,038**	-,145**	-,531**	,473**	,016*	1	-,069**
23 Prin. Instr. Lead.	,141**	-,033**	,138**	-,052**	,094**	,024**	,336**	,061**	-,095**	-.005	,033**	,015*	,066**	,193**	,092**	,231**	,123**	,093**	,059**	-,089**	,070**	-,069**	1

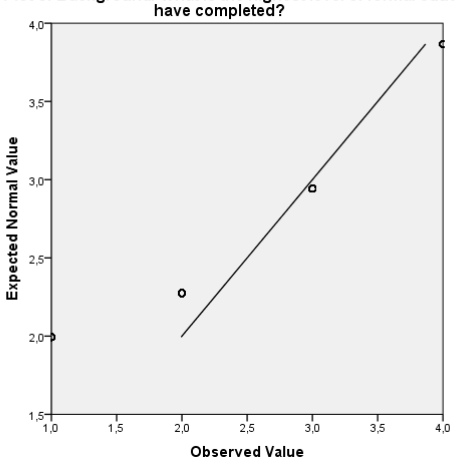
**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

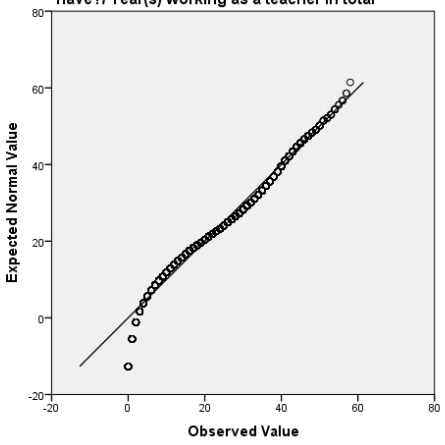
Appendix 5: QQ-plots for teacher level variables

The following graphs show QQplots for the level 1 variables (teacher and classroom). As can be seen from the plots below, most variables fit the normal line very well and thus homoscedasticity can be assumed; except the hours teachers spend on teaching and planning of lessons.

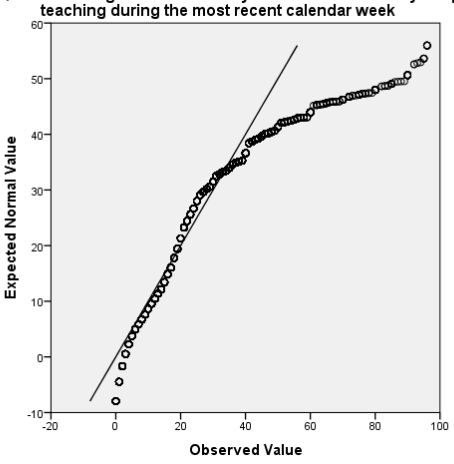
Normal Q-Q Plot of Background/ What is the highest level of formal education you have completed?



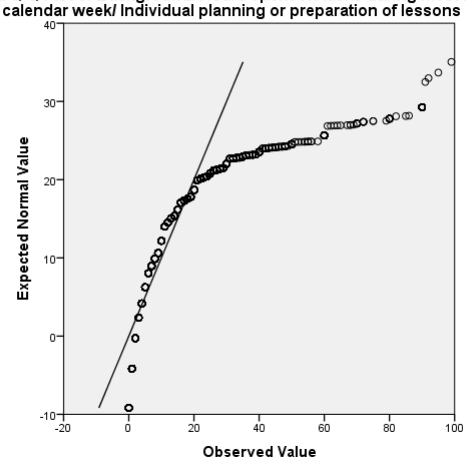
Normal Q-Q Plot of Background/ How many years of work experience do you have?/ Year(s) working as a teacher in total



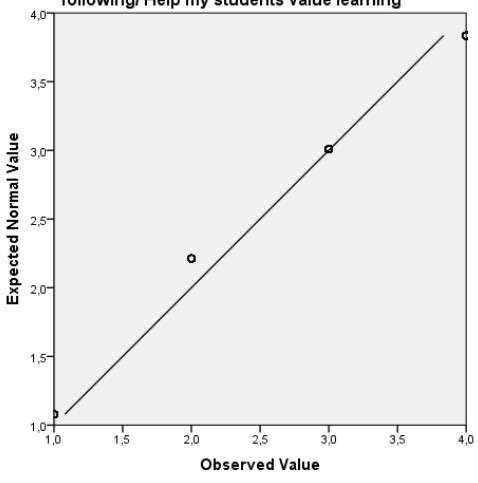
Normal Q-Q Plot of Background/ How many 60-minute hours did you spend on teaching during the most recent calendar week



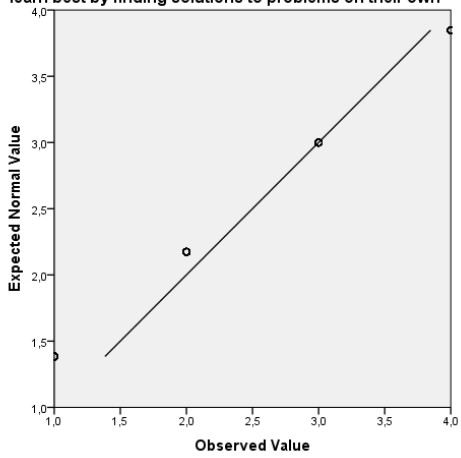
Normal Q-Q Plot of Background/ Hours spent on tasks during most recent calendar week/ Individual planning or preparation of lessons



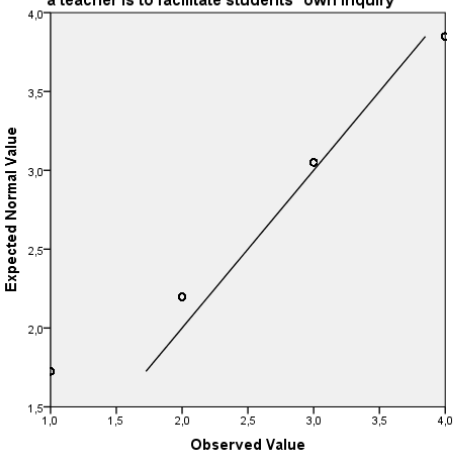
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Help my students value learning



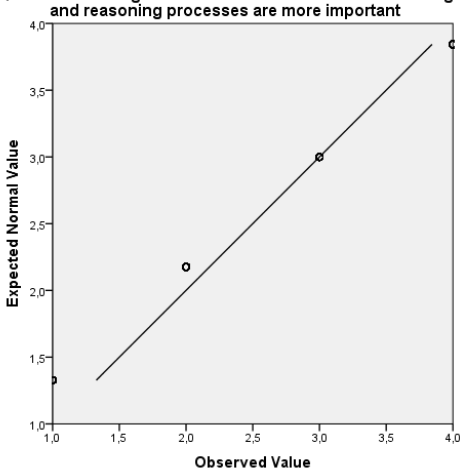
Normal Q-Q Plot of Teaching in General/ Personal beliefs on teaching/ Students learn best by finding solutions to problems on their own



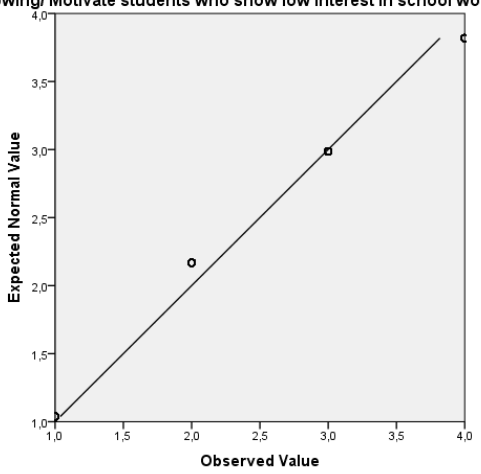
Normal Q-Q Plot of Teaching in General/ Personal beliefs on teaching/ My rol
a teacher is to facilitate students' own inquiry



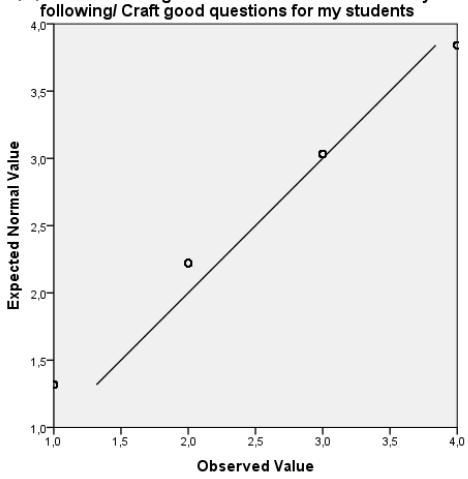
Normal Q-Q Plot of Teaching in General/ Personal beliefs on teaching/ Thinking and reasoning processes are more important



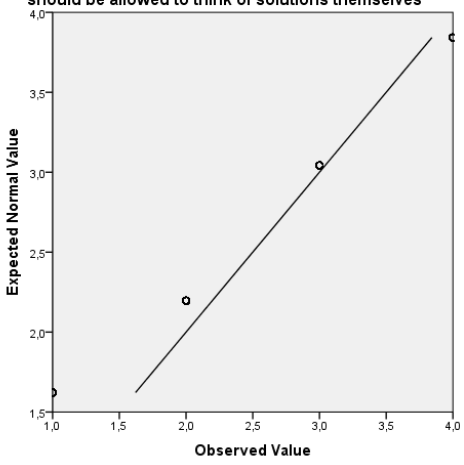
Normal Q-Q Plot of Teaching in General/ To what extend can you do tl
following/ Motivate students who show low interest in school work



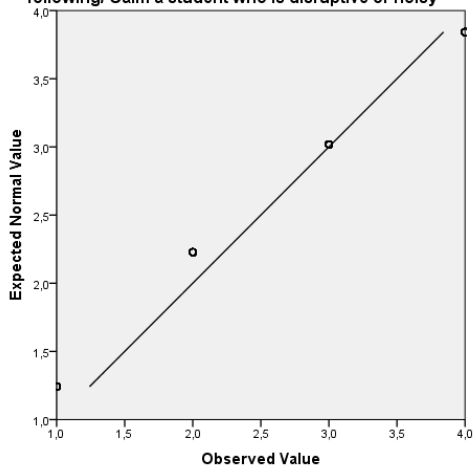
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Craft good questions for my students



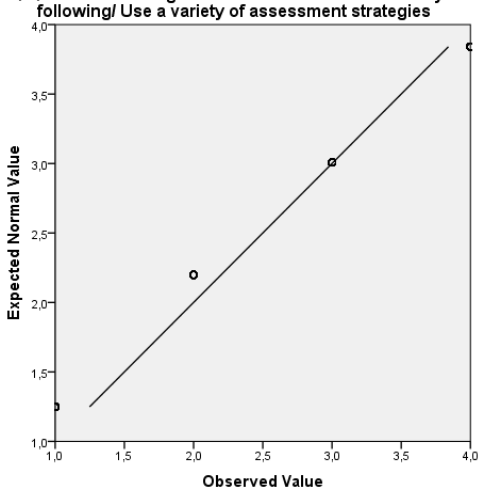
Normal Q-Q Plot of Teaching in General/ Personal beliefs on teaching/ Studen
should be allowed to think of solutions themselves



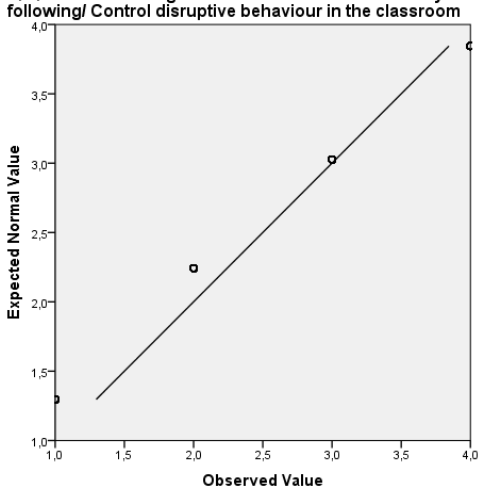
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Calm a student who is disruptive or noisy



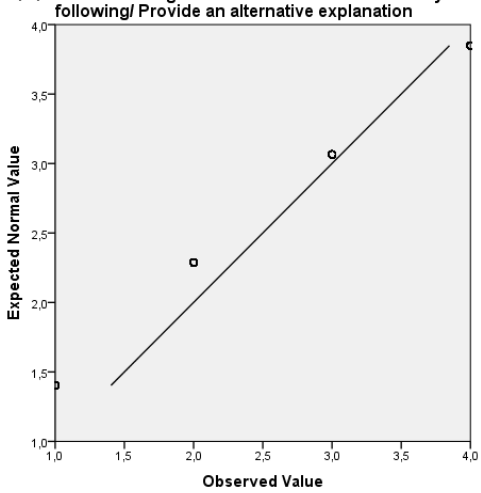
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Use a variety of assessment strategies



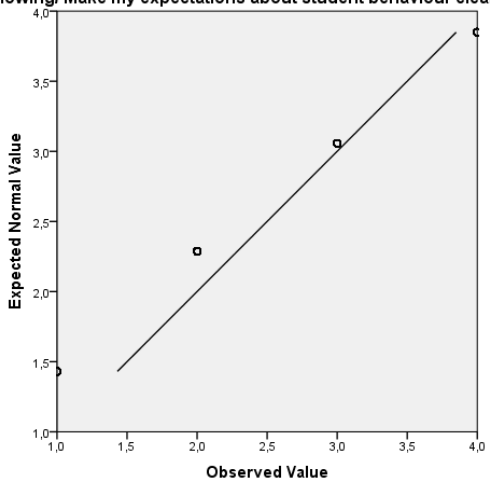
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Control disruptive behaviour in the classroom



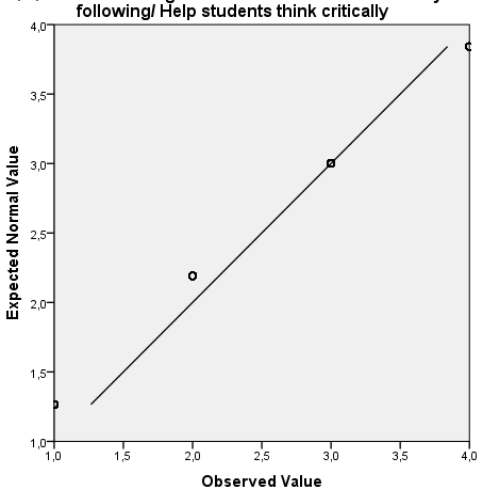
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Provide an alternative explanation



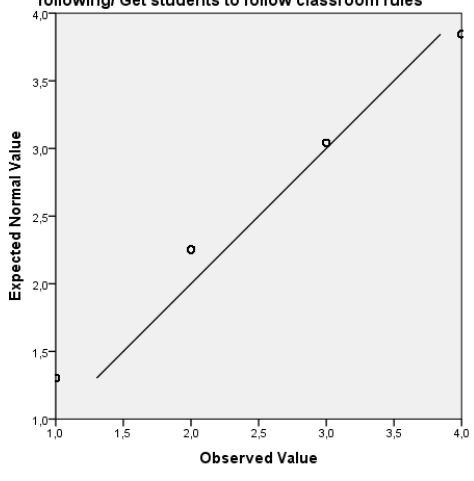
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Make my expectations about student behaviour clear



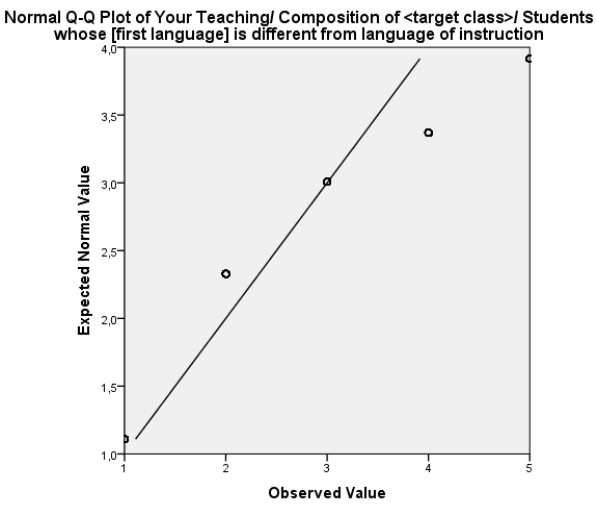
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Help students think critically

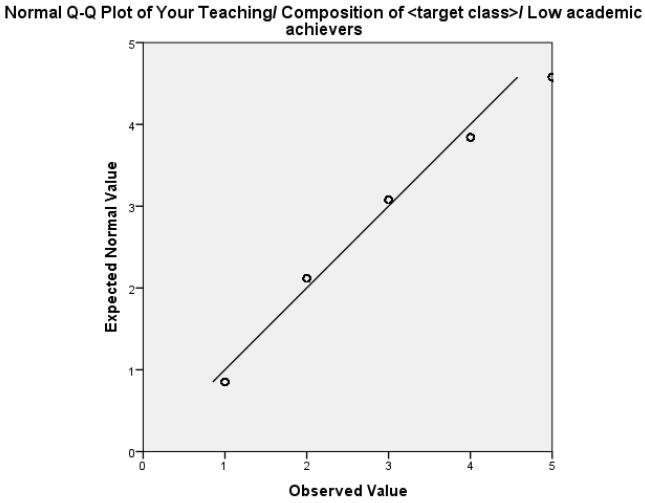


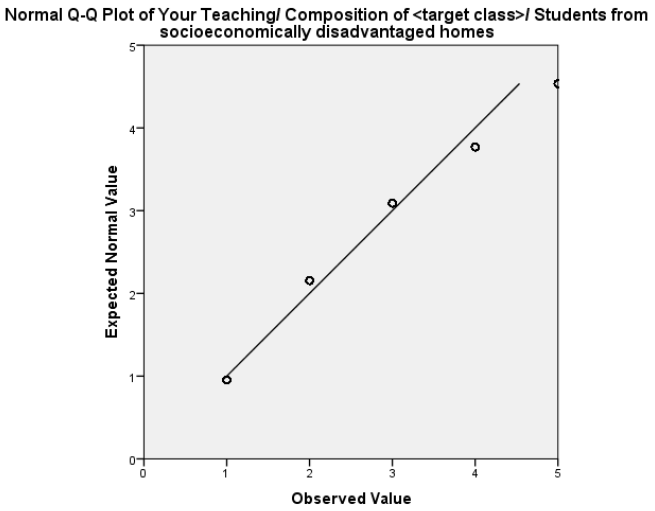
Normal Q-Q Plot of Teaching in General/ To what extend can you do the following/ Get students to follow classroom rules

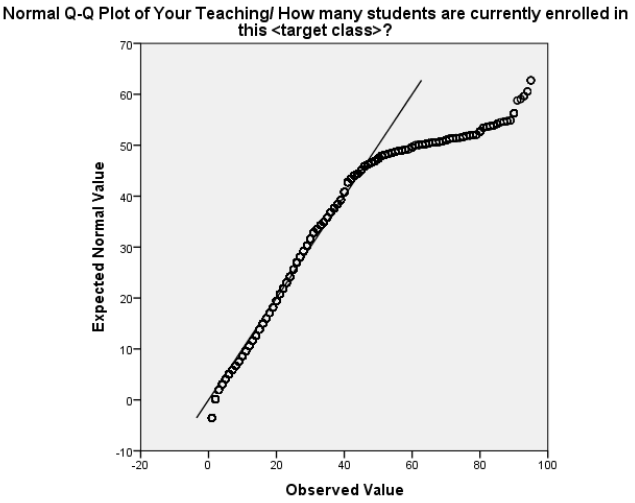


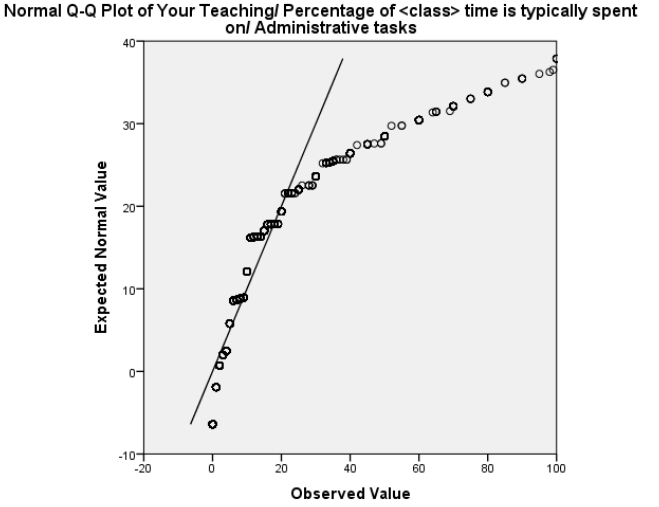
The classroom variables show a similar behaviour as the teacher variables: Those variables, which had deviating values for the skewness and kurtosis, also slightly deviate from the normal line in the QQ plots. Accordingly, homoscedasticity can be assumed for most of the variables.

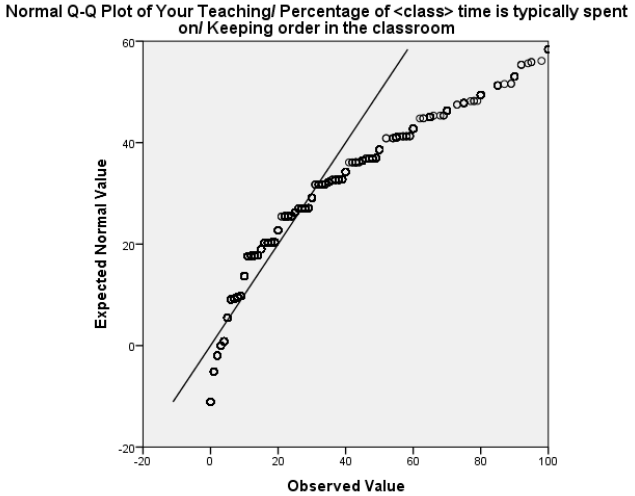




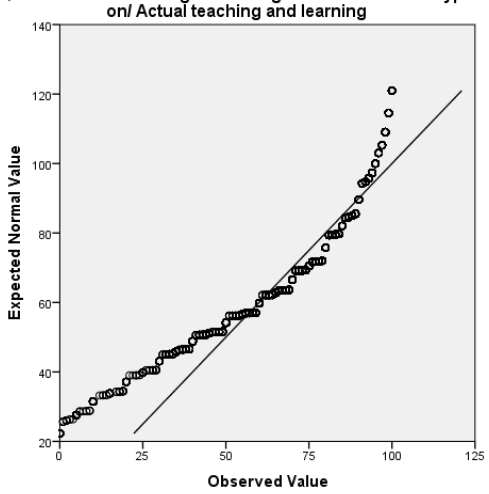




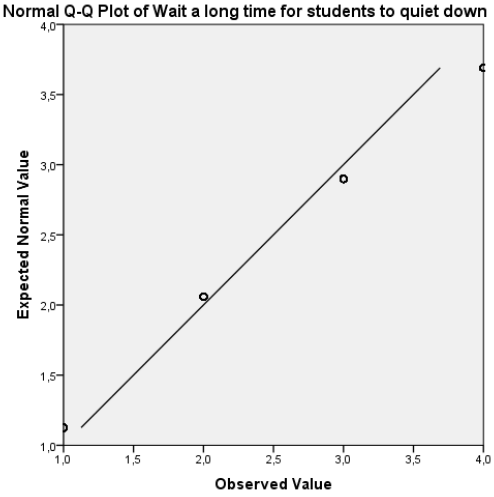




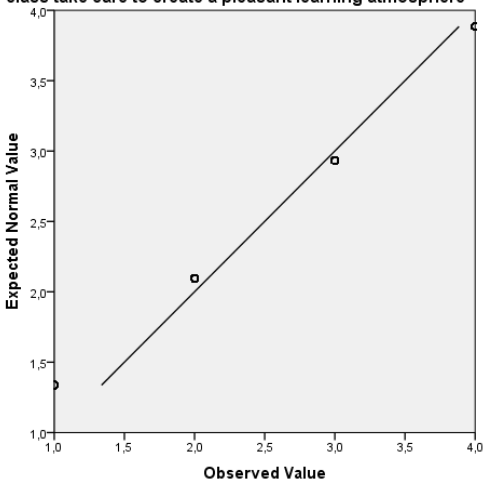
Normal Q-Q Plot of Your Teaching/ Percentage of <class> time is typically spent on/ Actual teaching and learning



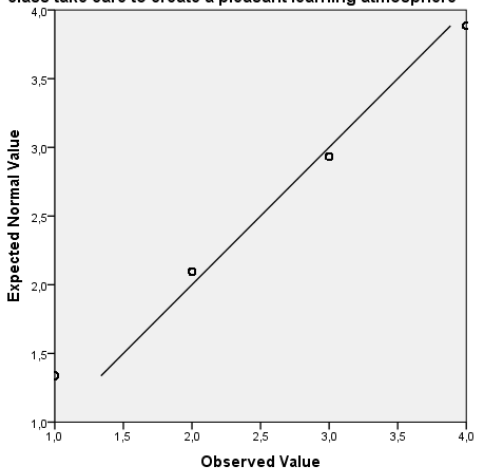
Normal Q-Q Plot of Wait a long time for students to quiet down



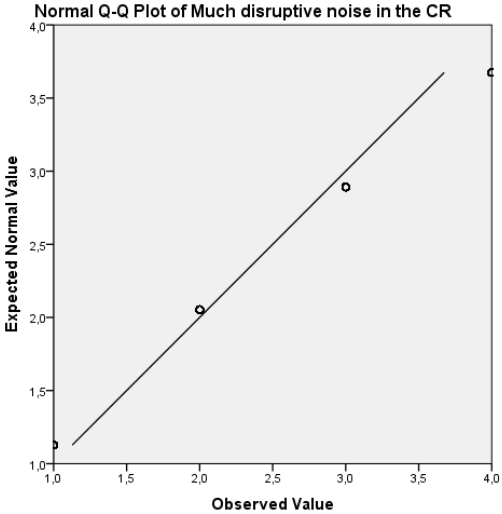
Normal Q-Q Plot of Your Teaching/ Agreement with statements/ Students in this class take care to create a pleasant learning atmosphere



Normal Q-Q Plot of Your Teaching/ Agreement with statements/ Students in this class take care to create a pleasant learning atmosphere



Normal Q-Q Plot of Much disruptive noise in the CR



Appendix 6: Significance tests for the predictors investigated

A different method of difference testing would have been to report whether the strength of the association between each sub-scale of TSE and each predictor between the three country clusters is significant or not. However, it was decided to report and compare the effect sizes instead, as they provide a measure, which is independent of sample sizes and were thus deemed superior. Results of the analysis will be reported below to show a different analytical approach and might be used for future research.

Similarities and differences in the country clusters for teacher's self-efficacy in instruction (ITSE)

This section presents the differential results for predictors of teachers' self-efficacy in instruction. It is divided into four parts, describing the differences for teacher-, classroom-principal- and school characteristics separately.

Teacher characteristics

At first, the difference in teacher characteristics will be reported. Part of the analysis were teacher gender, educational level, work experience, hours per week spent on teaching, planning lessons or administrative duties and the constructivist beliefs about teaching a teacher holds.

For female teacher gender, the difference between the Anglo Saxon and Nordic cluster did not appear to be significant, but the difference between the Anglo-Saxon and East- and South-East Asian was ($\boldsymbol{\Delta}_{\text{A-E}} = .042$, $p < .01$), as well as for the Nordic and East- and South-East Asian cluster ($\boldsymbol{\Delta}_{\text{N-E}} = .040$, $p < .01$). This means that being a female teacher in the Anglo-Saxon or Nordic countries predicted a higher level of ITSE than being a female teacher in an East- or South Asian country.

None of the differences occurring in the responses regarding the teachers' educational level and its association with ITSE in either of the cluster appeared to be statistically significant.

For the years of work experience, in contrast, all three differences were significant: Between the Anglo-Saxon and Nordic cluster ($\boldsymbol{\Delta}_{A\text{-}N}$ = .001, p < .05) it was positive, whereas between the Anglo-Saxon and East- and South-East Asian cluster the difference appeared to be very weak but negative ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.002, p < .01) as well as between the Nordic and the East- and South-East Asian cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.003, p < .001). Thus, both times in the ESEA cluster, the regression coefficient was slightly stronger than in the Anglo-Saxon or the Nordic countries.

The difference of the association between ITSE and the hours per week spent on teaching was significant between the Anglo-Saxon and Nordic cluster ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.003, p < .001) as well as for the Nordic and East- and South-East Asian cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = .002, p < .01). Accordingly, in the Nordic countries, the relationship was a little stronger than in the Anglo-Saxon countries and in the Nordic than in the East- and South-East Asian countries. The difference between Anglo-Saxon and East- and South-East Asian countries was not significant.

Between Anglo-Saxon and Nordic countries, there was found to be no differences in the magnitude of hours spent on planning lessons. Between Anglo-Saxon and East-and South-East Asian countries ($\boldsymbol{\Delta}_{A\text{-}E}$ = .002, p < .05) and Nordic and East- and South-East Asian countries ($\boldsymbol{\Delta}_{N\text{-}E}$ = .002, p < .05) respectively, there was a positive difference. Thus, putting more hours of lesson planning predicted a higher ITSE in Anglo-Saxon and Nordic countries than in East/South Asian ones.

For hours per week spent on administrative work, the picture is the same: There is no significant difference between Anglo-Saxon and Nordic countries. In contrast, the difference between Anglo-Saxon and East-and South-East Asian countries ($\boldsymbol{\Delta}_{A\text{-}E}$ = .003, p < .01) and Nordic and East- and South-East Asian countries ($\boldsymbol{\Delta}_{N\text{-}E}$ = .004, p < .01) respectively was found to be significant indicating that putting more hours of admin predicted a higher ITSE in Anglo-Saxon and Nordic countries than in East/South Asian ones.

Investigating the difference between the constructivist beliefs about teaching revealed that the differences between all three clusters were significant: The difference between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.007, p < .01) was negative, as was the difference

between Anglo-Saxon and East- and South-East Asian countries ($\Delta_{A\text{-}E}$ = -.018, p < .001) and Nordic and East- and South-East Asian countries ($\Delta_{N\text{-}E}$ = -.011, p < .01).

Classroom characteristics

This section reports and analyses the differences found in the responses for classroom characteristics. Subject of the analysis were class size, classroom climate, % of low achievers in the class, % of classroom time spent on administrative duties, keeping order and teaching and learning, the % of children with a different mother tongue than the language of instruction and % of children from disadvantaged homes.

Testing for the difference between class size showed that between Anglo-Saxon and Nordic countries ($\Delta_{A\text{-}E}$ = -.003, p < .001), there is a negative difference indicating that the association between ITSE and class size is significantly stronger in ESEA than in the Anglo-Saxon countries. The difference between Nordic and East- and South-East Asian countries ($\Delta_{N\text{-}E}$ = .003, p < .001), in contrast, was found to be positive with a stronger association in the Nordic countries. The difference between the Anglo-Saxon and Nordic countries was not significant.

None of the associations of the classroom climate with ITSE in the three country clusters differed significantly from each other.

For the percentage of classroom time spent on keeping order, the only significant effect was found for the difference between Nordic and East- and South-East Asian countries ($\Delta_{N\text{-}E}$ = -.003, p < .01).

Lastly, as with the time spent on teaching and learning, there was no significant difference found for the time spent on teaching and learning in Anglo-Saxon and Nordic countries and their association with ITSE. Between Anglo-Saxon and ESEA countries ($\Delta_{A\text{-}E}$ = -.003, p < .05), then, there was a negative difference as well as the Nordic and ESEA countries ($\Delta_{N\text{-}E}$ = -.003, p < .01).

Table 36 Differential results for ITSE

ITSE					
	Cluster	Variable	Δ-β	P	S.E.
Teacher					
	A-N	Female Gender	.002	n.s.	.013
	A-E	Female Gender	.042	**	.013
	N-E	Female Gender	.040	**	.012
	A-N	Edu. Level	.020	n.s.	.027
	A-E	Edu. Level	-.002	n.s.	.029
	N-E	Edu. Level	-.022	n.s.	.018
	A-N	Teaching Exp.	.001	*	.001
	A-E	Teaching Exp.	-.002	**	.001
	N-E	Teaching Exp.	-.003	***	.001
	A-N	Hrs. Teaching	-.003	**	.001
	A-E	Hrs. Teaching	-.001	n.s.	.001
	N-E	Hrs. Teaching	.002	**	.001
	A-N	Hrs. Planning	.000	n.s.	.001
	A-E	Hrs. Planning	.002	*	.001
	N-E	Hrs. Planning	.002	*	.001
	A-N	Hrs. Admin	-.001	n.s.	.002
	A-E	Hrs. Admin	.003	**	.001
	N-E	Hrs. Admin	.004	**	.001
	A-N	Construct. Bel.	-.007	**	.004
	A-E	Construct. Bel.	-.018	***	.003
	N-E	Construct. Bel.	-.011	**	.003
Classroom					
	A-N	Class size	-.003	***	.001
	A-E	Class size	.000	n.s.	.001
	N-E	Class size	.003	***	.001
	A-N	Classroom Climate	.001	n.s.	.003
	A-E	Classroom Climate	.001	n.s.	.003
	N-E	Classroom Climate	.000	n.s.	.003
	A-N	Low Student Achievement	-.011	n.s.	.008
	A-E	Low Student Achievement	.020	**	.008
	N-E	Low Student Achievement	.031	***	.007
	A-N	Teaching & Learning	.001	n.s.	.002
	A-E	Teaching & Learning	-.003	*	.002
	N-E	Teaching & Learning	-.003	**	.001
	A-N	Keeping Order	.001	n.s.	.002
	A-E	Keeping Order	-.002	n.s.	.002
	N-E	Keeping Order	-.003	**	.001
	A-N	Admin. Tasks	-.002	n.s.	.002
	A-E	Admin. Tasks	-.008	***	.002
	N-E	Admin. Tasks	-.006	***	.002
	A-N	Different Language	-.020	**	.007
	A-E	Different Language	-.022	**	.007
	N-E	Different Language	-.002	n.s.	.007
	A-N	Disadv. Background	.004	n.s.	.008
	A-E	Disadv. Background	-.025	**	.008
	N-E	Disadv. Background	-.029	***	.008

ITSE					
	Cluster	Variable	Δ-β	P	S.E.
Principal					
	A-N	Female Gender	.005	n.s.	.019
	A-E	Female Gender	-.074	**	.022
	N-E	Female Gender	-.079	***	.021
	A-N	Age	.009	***	.002
	A-E	Age	.009	***	.002
	N-E	Age	.011	***	.002
	A-N	Educational Level	.001	n.s.	.002
	A-E	Educational Level	-.005	**	.002
	N-E	Educational Level	-.005	**	.002
	A-N	Work Experience	-.002	n.s.	.001
	A-E	Work Experience	.009	***	.002
	N-E	Work Experience	.011	***	.002
	A-N	Instructional Leadership	.025	n.s.	.032
	A-E	Instructional Leadership	-.025	n.s.	.044
	N-E	Instructional Leadership	-.050	n.s.	.048
School					
	A-N	Student enrolment	.089	***	.024
	A-E	Student enrolment	.027	n.s.	.031
	N-E	Student enrolment	-.061	*	.035
	A-N	Private School	.018	**	.008
	A-E	Private School	.051	***	.008
	N-E	Private School	.032	***	.008
	A-N	School location	-.005	n.s.	.005
	A-E	School location	-.045	***	.005
	N-E	School location	-.040	***	.005

Investigating the difference of the association of ITSE and the percentage of students with a different mother tongue from the language of instruction in the classroom between Anglo-Saxon and Nordic countries (Δ_{A-N} = -.020, p < .01) revealed a negative difference. Between Anglo-Saxon and ESEA countries (Δ_{A-E} = -.022, p < .01), the difference was also significantly negative indicating that both times, the association in the Anglo-Saxon countries was weaker than in the Nordic or ESEA countries. The difference between Nordic and ESEA countries was not significant.

The difference between Anglo-Saxon and Nordic countries for the percentage of students from as disadvantaged background was not found to be significant. But, both differences between Anglo-Saxon and ESEA (Δ_{A-E} = -.025, p < .01) and Nordic and ESEA countries (Δ_{N-E} = -.029, p < .001) respectively were significant and negative. Accordingly, the relationship between ITSE and the percentage of students from a disadvantaged background was higher in the ESEA countries than in both the western country clusters.

Principal characteristics

In the following, the differences between the country clusters for the principal characteristics will be reported. The characteristics focused on are principal age, work experience, educational level and instructional leadership style practised at school.

For the principal's female gender, no difference was between the Anglo-Saxon and Nordic country clusters. The difference between the Anglo-Saxon and ESEA cluster (Δ_{A-E} = -.074, p < .01) and between the Nordic and ESEA cluster (Δ_{N-E} = -.079, p < .001), in contrast, was found to be both significant and negative. This means, the association in the ESEA countries is stronger than in the western countries.

A similar pattern arises for the principal's work experience. There is again no significant difference between the Anglo-Saxon and Nordic clusters, but a significant positive difference the Anglo-Saxon and ESEA clusters (Δ_{A-E} = .009, p < .001) and the Nordic and ESEA cluster (Δ_{N-E} = .011, p < .001). Here, the association between the principal's work experience and ITSE in the ESEA countries is weaker than in the Anglo-Saxon and Nordic countries.

Similarly, there was no significant difference found for the principal's educational level and its relationship with ITSE between the Anglo-Saxon and Nordic country cluster. The difference between the Anglo-Saxon and ESEA countries ($\Delta_{A\text{-}E}$ = -.005, p < .01) was negative, as was the difference between the Nordic and the ESEA countries ($\Delta_{N\text{-}E}$ = -.005, p < .01). Thus, the association in the East- and South-East countries was again bigger than in the western countries.

For the association between the principal's instructional leadership style and ITSE, there was no significant difference between all three country clusters

School characteristics

This sub-section reports the difference in the association in ITSE and the school's location, the status as a private school and the number of students enrolled at school between the three country clusters of Anglo-Saxon, Nordic and ESEA countries.

As this was already the case for the difference of many of the principal characteristics and ITSE, there was no significant difference found between the Anglo-Saxon and Nordic countries for the school's location. The difference between Anglo-Saxon and ESEA countries ($\Delta_{A\text{-}E}$ = -.045, p < .001) and Nordic and ESEA countries ($\Delta_{N\text{-}E}$ = -.040, p < .001) both was found to be negative indicating that the association in the ESEA countries is stronger than in the western countries.

For private schools, all three differences in its association with ITSE between the clusters were significant: A small difference appeared between the Anglo-Saxon and Nordic cluster ($\Delta_{A\text{-}N}$ = .018, p < .01). Between the Anglo-Saxon and the ESEA cluster ($\Delta_{A\text{-}E}$ = .051, p < .001), respectively the Nordic and ESEA cluster ($\Delta_{N\text{-}E}$ = .032, p < .001), the difference was moderate and positive.

The last school characteristic investigated was student enrolment. A significantly positive difference appeared between the Anglo-Saxon and Nordic countries ($\Delta_{A\text{-}N}$ = .089, p < .001) and significantly negative difference between the Nordic and ESEA countries ($\Delta_{N\text{-}E}$ = -.061, p < .05). Between the Anglo-Saxon and ESEA countries, the difference was not found to be significant.

Similarities and differences between the country clusters for teacher's self-efficacy in Student Engagement (ETSE*)*

The following sub-section presents the differential results for teacher, classroom, principal and school characteristics across the three country clusters teachers' self-efficacy in student engagement. Indices denote in which way the difference was calculated (i.e. $\boldsymbol{\Delta}_{A\text{-}E}$ showing the difference between the Anglo-Saxon and East- & South-East Asian cluster).

Teacher characteristics

The teacher characteristics investigated were female vs. male gender, educational level, years of teaching experience, weekly hours spent on teaching, preparing lessons or administrative work and the degree to which a teacher holds constructivist beliefs about teaching.

For teacher's self-efficacy in student engagement, neither female gender nor the educational level revealed significant differences between the three country clusters.

Regarding years of work experience, there was no significant differences found between Anglo-Saxon and Nordic teachers. Between Anglo-Saxon and ESEA teachers ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.004, $p < .001$) and Nordic and ESEA teachers ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.004, $p < .001$), the difference was negative. The negative difference indicates that the association between years of work experience and ETSE in the ESEA countries is bigger than in the Anglo-Saxon and Nordic countries.

Comparing the three types of work allocation, the association for the hours per week spent on teaching revealed a significant difference between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.004, $p < .001$) and Nordic and ESEA countries ($\boldsymbol{\Delta}_{N\text{-}E}$ = .005, $p < .001$), but not between Anglo-Saxon and Nordic countries. For hours spent on planning lessons, there was no significant difference between all three clusters. Finally, there only appeared to be a significant difference between Anglo-Saxon and ESEA countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = .004, $p < .01$) for hours spent on administrative duties, but not between Anglo-Saxon and Nordic or Nordic and ESEA countries.

Table 37 Differential results for ETSE

ETSE					
	Cluster	Variable	Δ-β	P	S.E.
Teacher					
	A-N	Female Gender	.007	n.s.	.017
	A-E	Female Gender	.020	n.s.	.017
	N-E	Female Gender	.014	n.s.	.014
	A-N	Edu. Level	-.037	n.s.	.039
	A-E	Edu. Level	-.031	n.s.	.039
	N-E	Edu. Level	.006	n.s.	.022
	A-N	Teaching Exp.	.001	n.s.	.001
	A-E	Teaching Exp.	-.004	***	.001
	N-E	Teaching Exp.	-.004	***	.001
	A-N	Hrs. Teaching	-.004	***	.001
	A-E	Hrs. Teaching	.001	n.s.	.001
	N-E	Hrs. Teaching	.005	***	.001
	A-N	Hrs. Planning	.001	n.s.	.002
	A-E	Hrs. Planning	.002	n.s.	.001
	N-E	Hrs. Planning	.002	n.s.	.001
	A-N	Hrs. Admin	.003	n.s.	.002
	A-E	Hrs. Admin	.004	**	.002
	N-E	Hrs. Admin	.001	n.s.	.002
	A-N	Construct. Bel.	-.007	n.s.	.005
	A-E	Construct. Bel.	-.012	**	.004
	N-E	Construct. Bel.	-.005	n.s.	.004
Classroom					
	A-N	Class size	-.002	**	.001
	A-E	Class size	.000	n.s.	.001
	N-E	Class size	.002	**	.001
	A-N	Classroom Climate	.013	**	.004
	A-E	Classroom Climate	.017	***	.004
	N-E	Classroom Climate	.004	n.s.	.003
	A-N	Low Student Achievement	-.004	n.s.	.011
	A-E	Low Student Achievement	.022	**	.010
	N-E	Low Student Achievement	.026	**	.009
	A-N	Teaching & Learning	-.001	n.s.	.003
	A-E	Teaching & Learning	-.003	n.s.	.002
	N-E	Teaching & Learning	-.001	n.s.	.002
	A-N	Keeping Order	-.003	n.s.	.003
	A-E	Keeping Order	-.003	n.s.	.002
	N-E	Keeping Order	.000	n.s.	.002
	A-N	Admin. Tasks	-.002	n.s.	.003
	A-E	Admin. Tasks	-.007	**	.002
	N-E	Admin. Tasks	-.005	**	.002
	A-N	Different Language	.009	n.s.	.009
	A-E	Different Language	-.010	n.s.	.009
	N-E	Different Language	-.019	**	.009
	A-N	Disadv. Background	.002	n.s.	.011
	A-E	Disadv. Background	-.021	n.s.	.011
	N-E	Disadv. Background	-.023	**	.010

ETSE					
	Cluster	Variable	Δ-β	P	S.E.
Principal					
	A-N	Female Gender	.036	n.s.	.029
	A-E	Female Gender	-.117	**	.034
	N-E	Female Gender	-.153	***	.034
	A-N	Age	.016	***	.003
	A-E	Age	.016	***	.003
	N-E	Age	.017	***	.003
	A-N	Educational Level	-.006	**	.002
	A-E	Educational Level	-.011	**	.004
	N-E	Educational Level	-.004	n.s.	.003
	A-N	Work Experience	-.001	n.s.	.002
	A-E	Work Experience	.016	***	.003
	N-E	Work Experience	.017	***	.003
	A-N	Instructional Leadership	.004	n.s.	.046
	A-E	Instructional Leadership	-.041	n.s.	.076
	N-E	Instructional Leadership	-.045	n.s.	.077
School					
	A-N	Student enrolment	.143	***	.035
	A-E	Student enrolment	.023	n.s.	.053
	N-E	Student enrolment	-.121	**	.055
	A-N	Private School	.032	**	.011
	A-E	Private School	.091	***	.013
	N-E	Private School	.059	***	.013
	A-N	School location	.008	n.s.	.007
	A-E	School location	-.086	***	.009
	N-E	School location	-.094	***	.008

Similarly, for constructivist beliefs about teaching the only significant difference was found between Anglo-Saxon and ESEA teachers ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.012, p < .01), indicating that the association with ETSE for ESEA teachers was stronger than for Anglo-Saxon teachers.

Classroom characteristics

In total, the differences in eight classroom characteristics were analysed. These were the class size, classroom climate, percentage of low achievers in the class, percentage of classroom time spent on teaching and learning, keeping order or administrative duties, the percentage of pupils with a different mother tongue than the language of instruction and from a disadvantaged home.

Between Anglo-Saxon and Nordic teachers ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.002, p < .01), there was a significant negative difference for the class size and its relation to ETSE. Thus, the association was stronger for the Nordic teachers. Also, a significant but positive difference was found between Nordic and ESEA teachers ($\boldsymbol{\Delta}_{N\text{-}E}$ = .002, p < .01) with a stronger relation in the Nordic countries. The difference between Anglo-Saxon and ESEA countries, tough, appeared not to be significant.

The classroom climate revealed a significant positive difference between the Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = .013, p < .01) and a significant positive difference between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = .017, p < .001). Between Nordic and ESEA countries, however, there was no significant difference in the magnitude of the association between classroom climate and ETSE.

The percentage of low achievers in the classroom revealed a significant positive difference between Anglo-Saxon and ESEA teachers ($\boldsymbol{\Delta}_{A\text{-}E}$ = .022, p < .01) and Nordic and ESEA teachers ($\boldsymbol{\Delta}_{N\text{-}E}$ = .026, p < .01). Both times, the association in the western countries was stronger than in the eastern countries. The difference between Anglo-Saxon and Nordic teachers was not significant.

Comparing the three types of time allocation revealed only a significant difference for the percentage of time spent on administrative duties between Anglo-Saxon and ESEA teachers ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.007, p < .01) and Nordic and ESEA teachers ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.005, p < .01), but not between Anglo-Saxon and Nordic teachers. For the percentage of time spent on keeping order in the

classroom and spent of teaching and learning and its association with ETSE there was no difference found between any of the country clusters.

Investigating the relationship for the percentage of students in the classroom with a different mother tongue than the language of instruction showed only a significant difference between the Nordic and ESEA countries ($\boldsymbol{\Delta}_{\text{N-E}}$ = -.019, p < .01). A similar pattern was found for the percentage of students from disadvantaged backgrounds in the classroom, where there was only a significant difference between teaches in the Nordic and ESEA cluster. ($\boldsymbol{\Delta}_{\text{N-E}}$ = -.023, p < .01).

Principal characteristics

Relevant for the analysis were five principal characteristics: Female vs. male gender, age, the educational level, years of work experience and the degree, to which the principal employs an instructional leadership style.

Investigating the differences in the principal characteristics and their relationship with ETSE revealed no significant difference for the principal's female gender between the Anglo-Saxon and Nordic countries, but between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{\text{A-E}}$ = -.117, p < .001) and the Nordic and the ESEA cluster ($\boldsymbol{\Delta}_{\text{N-E}}$ = -.153, p < .001). For the principal's age there was a significant difference between the Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{\text{A-N}}$ = .016, p < .001), and between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{\text{A-E}}$ = .016, p < .001) and the Nordic and ESEA cluster ($\boldsymbol{\Delta}_{\text{N-E}}$ = .017, p < .001). Accordingly, the magnitude of the association is stronger in the ESEA than in the Anglo-Saxon or Nordic cluster. Similarly, there was no significant difference found for the Anglo-Saxon and Nordic countries with respect to the principal's work experience. However, the difference between Anglo-Saxon and ESEA countries ($\boldsymbol{\Delta}_{\text{A-E}}$ = .016, p < .001) and Nordic and ESEA countries ($\boldsymbol{\Delta}_{\text{N-E}}$ = .017, p < .001) was found to be significant and positive. Thus, the magnitude of the association between ETSE and the principal's work experience is bigger this time in the western than the eastern countries.

There is a different picture for the principal's educational level. Here, a significant difference occurs between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{\text{A-N}}$ = -.006, p < .01) and Anglo-Saxon and

ESEA countries (Δ_{A-E} = -.011, p < .01). Between Nordic and ESEA countries, the difference was not significant.

Finally, no significant differences were found between any of the three country clusters for the difference in magnitude of the association between the principal's instructional leadership style and ETSE.

School characteristics

Finally, there were three school characteristics in the model: Number of students enrolled at school, whether it is private or public school and the size of the community, in which the school is located. Investigating the differences for these school characteristics and the difference of their association with ETSE across the three different country clusters found multiple significant differential results.

For the school's location there was no significant difference found for Anglo-Saxon and Nordic countries. Between Anglo-Saxon and ESEA countries (Δ_{A-E} = -.086, p < .001), however, and Nordic and ESEA countries (Δ_{N-E} = -.094, p < .001) the difference was significant indicating that in the eastern cluster the association was stronger than in the western clusters.

For private schools, all three clusters showed significant differences in their association with ETSE. Between the Anglo-Saxon and Nordic cluster (Δ_{A-N} = .032, p < .01) the difference was positive indicating that the association in the Anglo-Saxon countries was stronger. Between the Anglo-Saxon and ESEA cluster (Δ_{A-E} = .091, p < .001) and the Nordic and ESEA cluster (Δ_{N-E} = .059, p < .001), the difference was also positive and highly significant.

Lastly, for the numbers of students enrolled at the respective school, the difference was positive between the Anglo-Saxon and Nordic cluster (Δ_{A-N} = .143, p < .001) and negative between the Nordic and ESEA cluster (Δ_{N-E} = -.121, p < .01). Between the Anglo-Saxon and ESEA cluster, however, the difference was not significant.

Similarities and differences in the country clusters for teacher's self-efficacy in Classroom Management (MTSE)

This sub-section presents the results of the difference in the magnitude of the association that MTSE has with each of the teacher-, classroom-, principal- and school characteristics. As with the previous sections, the $\boldsymbol{\Delta}_{A\text{-}E}$-notation describes the magnitude of the difference in the association with MTSE between two country clusters.

Teacher characteristics

The teacher characteristics analysed were female gender, educational level, years of teaching experience, weekly hours spent on teaching, planning lessons or administrative tasks and the degree of constructivist beliefs about teaching the teachers have.

Investigating the difference for teacher gender, revealed significant differences between all three country clusters: between the Anglo-Saxon and Nordic cluster there was positive difference ($\boldsymbol{\Delta}_{A\text{-}N}$ = .032, p < .05), as well as between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = .101, p < .001) and the Nordic and ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = .069, p < .001).

For the teacher's educational level, the difference was significant and negative between the Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.074, p < .05) and the Anglo-Saxon and ESEA countries ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.067, p < .05) indicating that the magnitude of the association in in ESEA and Nordic cluster was bigger than in the Anglo-Saxon cluster. Between the Nordic and the EASEA cluster, the difference was not found to be significant.

Similarly, a significant difference was found for the teacher's work experience between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.002, p < .01) and Anglo-Saxon and ESEA countries ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.002, p < .01), but not between Nordic and ESEA countries.

The picture for the work allocation per week was found to be very heterogenic. The difference in the hours per week spent on teaching was found to be significant between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.005, p < .001) and Nordic and ESEA countries ($\boldsymbol{\Delta}_{N\text{-}E}$ = 006, p < .001), but not between Anglo-Saxon and ESEA countries. The weekly hours spent on planning

lessons, then, were not significantly differently associated to MTSE between all three country clusters. For hours per week spent on administrative duties there was only a significant difference found between Anglo-Saxon and ESEA countries ($\boldsymbol{\Delta}_{A\text{-}E}$ = .004, p < .01), but not between the Anglo-Saxon and Nordic or Nordic and ESEA cluster.

Finally, the differences for constructivist beliefs about teaching and their association with MTSE in the three country clusters were all significant: between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.008, p < .05) the difference was negative, as it was between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.035, p < .001) and the Nordic and ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.026, p < .001).

Classroom characteristics

Regarding the classroom, the class size, the perceived classroom climate and the percentage of low achievers were investigated. Furthermore, time allocation in the classroom as to how much time was spent on teaching and learning, keeping order or administrative tasks as well as the percentage of students with a different mother tongue and the percentage of students from a disadvantaged home were analysed.

Investigating the difference for the classroom size between Anglo-Saxon and Nordic countries ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.004, p < .001) revealed a negative small, but highly significant difference. Between Nordic and ESEA countries ($\boldsymbol{\Delta}_{N\text{-}E}$ = .003, p < .01), the difference was small but positive, but no significant difference was found between the Anglo-Saxon and ESEA cluster.

With respect to the classroom climate and the magnitude of its association with MTSE, there was no significant difference found between any of the three country clusters.

The association for the percentage of low achievers in the classroom was not found to be different between the Anglo-Saxon and Nordic and the Anglo-Saxon and ESEA cluster. In contrast, the association between the Nordic and the ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = .018, p < .05) was positive and significant.

Table 38 Differential results for MTSE

MTSE

	Cluster	Variable	Δ-β	P	S.E.
Teacher					
	A-N	Female Gender	.032	*	.018
	A-E	Female Gender	.101	***	.018
	N-E	Female Gender	.069	***	.016
	A-N	Edu. Level	-.074	*	.038
	A-E	Edu. Level	-.067	*	.039
	N-E	Edu. Level	.008	n.s.	.025
	A-N	Teaching Exp.	-.002	**	.001
	A-E	Teaching Exp.	-.002	**	.001
	N-E	Teaching Exp.	.000	n.s.	.001
	A-N	Hrs. Teaching	-.005	***	.001
	A-E	Hrs. Teaching	.001	n.s.	.001
	N-E	Hrs. Teaching	.006	***	.001
	A-N	Hrs. Planning	-.002	n.s.	.002
	A-E	Hrs. Planning	-.002	n.s.	.001
	N-E	Hrs. Planning	.000	n.s.	.002
	A-N	Hrs. Admin	.003	n.s.	.002
	A-E	Hrs. Admin	.004	**	.002
	N-E	Hrs. Admin	.001	n.s.	.002
	A-N	Construct. Bel.	-.008	*	.005
	A-E	Construct. Bel.	-.035	***	.005
	N-E	Construct. Bel.	-.026	***	.004
Classroom					
	A-N	Class size	-.004	***	.001
	A-E	Class size	-.001	n.s.	.001
	N-E	Class size	.003	**	.001
	A-N	Classroom Climate	.002	n.s.	.005
	A-E	Classroom Climate	.007	n.s.	.005
	N-E	Classroom Climate	.004	n.s.	.005
	A-N	Low Student Achievement	-.005	n.s.	.011
	A-E	Low Student Achievement	.013	n.s.	.011
	N-E	Low Student Achievement	.018	*	.010
	A-N	Teaching & Learning	-.001	n.s.	.002
	A-E	Teaching & Learning	-.003	n.s.	.002
	N-E	Teaching & Learning	-.003	n.s.	.002
	A-N	Keeping Order	.001	n.s.	.002
	A-E	Keeping Order	-.005	**	.002
	N-E	Keeping Order	-.005	**	.002
	A-N	Admin. Tasks	-.003	n.s.	.003
	A-E	Admin. Tasks	-.007	**	.002
	N-E	Admin. Tasks	-.004	n.s.	.003
	A-N	Different Language	.005	n.s.	.010
	A-E	Different Language	-.028	**	.009
	N-E	Different Language	-.033	***	.009
	A-N	Disadv. Background	.004	n.s.	.011
	A-E	Disadv. Background	-.014	n.s.	.011
	N-E	Disadv. Background	-.019	*	.010

MTSE

	Cluster	Variable	Δ-β	P	S.E.
Principal					
	A-N	Female Gender	-.008	n.s.	.054
	A-E	Female Gender	-.133	***	.190
	N-E	Female Gender	-.125	***	.178
	A-N	Age	.009	***	.002
	A-E	Age	.009	***	.002
	N-E	Age	.011	***	.002
	A-N	Educational Level	-.005	**	.002
	A-E	Educational Level	-.010	**	.003
	N-E	Educational Level	-.005	*	.003
	A-N	Work Experience	-.003	n.s.	.002
	A-E	Work Experience	.009	***	.002
	N-E	Work Experience	.011	***	.002
	A-N	Instructional Leadership	-.017	n.s.	.049
	A-E	Instructional Leadership	.031	n.s.	.065
	N-E	Instructional Leadership	.048	n.s.	.066
School					
	A-N	Student enrolment	.113	***	.030
	A-E	Student enrolment	.054	n.s.	.044
	N-E	Student enrolment	-.059	n.s.	.046
	A-N	Private School	.010	n.s.	.011
	A-E	Private School	.090	***	.011
	N-E	Private School	.080	***	.011
	A-N	School location	.004	n.s.	.006
	A-E	School location	-.059	***	.007
	N-E	School location	-.062	***	.007

The time spent on different classroom activities drew a heterogenic picture: the percentage of time spent on administrative duties showed only a significant negative difference between Anglo-Saxon and ESEA countries ($\Delta_{A\text{-}E}$ = -.007, p < .01). The percentage of time spent on keeping order in the classroom only showed a significant difference between the Anglo-Saxon and ESEA countries ($\Delta_{A\text{-}E}$ = -.005, p < .01) and the Nordic and the ESEA countries ($\Delta_{N\text{-}E}$ = -.005, p < .01), but not between the Anglo-Saxon and Nordic cluster. Finally, the percentage of classroom time spent on teaching and learning and its association with MTSE revealed no significant differences between either country cluster.

For the percentage of students in the class who speak a different mother tongue than the language of instruction showed a difference between the Anglo-Saxon and ESEA country cluster ($\Delta_{A\text{-}E}$ = -.028, p < .01) and the Nordic and ESEA country cluster ($\Delta_{N\text{-}E}$ = -.033, p < .001), but not between the Anglo-Saxon and Nordic countries. Thus, the magnitude of the association was bigger in the ESEA cluster than in the Anglo-Saxon or Nordic cluster.

The percentage of students from socio-economically disadvantaged backgrounds only showed a significant difference between the Nordic and the ESEA cluster ($\Delta_{N\text{-}E}$ = -.019, p < .05). The other differences between the Anglo-Saxon and ESEA and the Anglo-Saxon and the Nordic cluster, in contrast, were not significant.

Principal characteristics

Concerning the principals, their gender, age, educational level, years of work experience as principals and degree of instructional leadership style were taken into account.

The principal's female gender revealed no significant difference in its relationship with MTSE between the Anglo-Saxon and the Nordic cluster. Between the Anglo-Saxon and ESEA cluster ($\Delta_{A\text{-}E}$ = -.133, p < .001) and the Nordic and ESEA cluster ($\Delta_{N\text{-}E}$ = -.125, p < .001), respectively, there was a comparatively strong negative difference. Thus, the magnitude of the association was stronger in the ESEA cluster than in the Anglo-Saxon

or Nordic cluster. For the principal's age, there were positive significant differences between all three clusters. Between the Anglo-Saxon and the Nordic cluster it was ($\boldsymbol{\Delta}_{A\text{-}N}$ = .009, p < .01), between the Anglo-Saxon and the ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = .009, p < .001) and between the Nordic and ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = .011, p < .001).

Similarly, there was no significant difference between the Anglo-Saxon and Nordic cluster with respect to the principal's work experience. Between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = .009, p < .001) and the Nordic and ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = .011, p < .001), however, the difference found was significant and positive.

For the educational level, all three differences were found to be significant and negative. Between the Anglo-Saxon and Nordic cluster ($\boldsymbol{\Delta}_{A\text{-}N}$ = -.005, p < .01) the difference was very small, and only slightly bigger between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.010, p < .01). The difference between the Nordic and the ESEA cluster was also found to be small ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.005, p < .05) again.

None of the country clusters had a different significance with another cluster with respect to the principal's instructional leadership style and its association with MTSE.

School characteristics

Finally, three school characteristics were compared: The number of students enrolled at school, the status as a private or public school and the size of the community in which the school is located. For those school characteristics, the associations with MTSE in the country clusters were again very heterogenic.

Between the Anglo-Saxon and Nordic cluster, no difference was found with respect to the school's location. But between the Anglo-Saxon and ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = -.059, p < .001) and the Nordic and the ESEA cluster ($\boldsymbol{\Delta}_{N\text{-}E}$ = -.062, p < .001), the differences found were moderately large and negative.

Similarly, there was no significant difference found between the Anglo-Saxon and the Nordic countries for the schools' private status and the magnitude of its relationship with MTSE. Between the Anglo-Saxon and the ESEA cluster ($\boldsymbol{\Delta}_{A\text{-}E}$ = .090, p < .001), in

contrast, the difference in the relationship was positive and moderately large, between the Nordic and the ESEA cluster ($\Delta_{N\text{-}E}$ = .080, p < .001). Thus, the association in the ESEA countries was smaller than in the Anglo-Saxon and Nordic countries.

Finally, the relationship between MTSE and student enrolment was only found to be significantly different in the Anglo-Saxon and Nordic countries ($\Delta_{A\text{-}N}$ = .113, p < .001). Between the other country clusters, the difference was insignificant.