

IEA Research for Education

A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)



Michalis P. Michaelides

Gavin T. L. Brown

Hanna Eklöf

Elena C. Papanastasiou

Motivational Profiles in TIMSS Mathematics

Exploring Student Clusters Across
Countries and Time



 Springer Open

IEA Research for Education

A Series of In-depth Analyses Based on Data
of the International Association for the Evaluation
of Educational Achievement (IEA)

Volume 7

Series Editors

Seamus Hegarty, University of Warwick, UK, and Chair of IEA Publications and Editorial Committee

Leslie Rutkowski, Indiana University, USA

Editorial Board

John Ainley, Australian Council for Educational Research, Australia

Kadriye Ercikan, University of British Columbia, Canada

Eckhard Klieme, German Institute for International Educational Research (DIPF), Germany

Rainer Lehmann, Humboldt University of Berlin, Germany

Fou-Lai Lin, National Taiwan Normal University, Chinese Taipei

Marlaine Lockheed, Princeton University, USA

Sarah Maughan, AlphaPlus Consultancy, UK

Carina Omoeva, FHI 360, USA

Elena C. Papanastasiou, University of Nicosia, Cyprus

Valena White Plisko, Independent Consultant, USA

Jonathan Plucker, John Hopkins University, USA

Fernando Reimers, Harvard Graduate School of Education, USA

David Rutkowski, Indiana University, USA

Jouni Välijärvi, University of Jyväskylä, Finland

Hans Wagemaker, Senior Advisor to IEA, New Zealand

The International Association for the Evaluation of Educational Achievement (IEA) is an independent nongovernmental nonprofit cooperative of national research institutions and governmental research agencies that originated in Hamburg, Germany in 1958. For over 60 years, IEA has developed and conducted high-quality, large-scale comparative studies in education to support countries' efforts to engage in national strategies for educational monitoring and improvement.

IEA continues to promote capacity building and knowledge sharing to foster innovation and quality in education, proudly uniting more than 60 member institutions, with studies conducted in more than 100 countries worldwide.

IEA's comprehensive data provide an unparalleled longitudinal resource for researchers, and this series of in-depth peer-reviewed thematic reports can be used to shed light on critical questions concerning educational policies and educational research. The goal is to encourage international dialogue focusing on policy matters and technical evaluation procedures. The resulting debate integrates powerful conceptual frameworks, comprehensive datasets, and rigorous analysis, thus enhancing understanding of diverse education systems worldwide.

More information about this series at <http://www.springer.com/series/14293>

Michalis P. Michaelides ·
Gavin T. L. Brown · Hanna Eklöf ·
Elena C. Papanastasiou

Motivational Profiles in TIMSS Mathematics

Exploring Student Clusters Across Countries
and Time



Michalis P. Michaelides
Department of Psychology
University of Cyprus
Nicosia, Cyprus

Hanna Eklöf
Umeå University
Umeå, Sweden

Gavin T. L. Brown
The University of Auckland
Auckland, New Zealand

Elena C. Papanastasiou
Department of Education
University of Nicosia
Nicosia, Cyprus



ISSN 2366-1631

ISSN 2366-164X (electronic)

IEA Research for Education

ISBN 978-3-030-26182-5

ISBN 978-3-030-26183-2 (eBook)

<https://doi.org/10.1007/978-3-030-26183-2>

© International Association for the Evaluation of Educational Achievement (IEA) 2019. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

This work is subject to copyright. All commercial rights are reserved by the author(s), whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Regarding these commercial rights a non-exclusive license has been granted to the publisher.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

IEA's mission is to enhance knowledge about education systems worldwide, and to provide high-quality data that will support education reform and lead to better teaching and learning in schools. In pursuit of this aim, it conducts, and reports on, major studies of student achievement in literacy, mathematics, science, citizenship, and digital literacy. These studies, most notably TIMSS, PIRLS, ICCS, and ICILS, are well established and have set the benchmark for international comparative studies in education.

The studies have generated vast datasets encompassing student achievement, disaggregated in a variety of ways, along with a wealth of contextual information which contains considerable explanatory power. The numerous reports that have emerged from them are a valuable contribution to the corpus of educational research.

Valuable though these detailed reports are, IEA's goal of supporting education reform needs something more: deep understanding of education systems and the many factors that bear on student learning advances through in-depth analysis of the global datasets. IEA has long championed such analysis and facilitates scholars and policymakers in conducting secondary analysis of our datasets. So, we provide software such as the International Database Analyzer to encourage the analysis of our datasets, support numerous publications including a peer-reviewed journal—*Large-scale Assessments in Education*—dedicated to the science of large-scale assessment and publishing articles that draw on large-scale assessment databases, and organize a biennial international research conference to nurture exchanges between researchers working with IEA data.

The **IEA Research for Education** series represents a further effort by IEA to capitalize on our unique datasets, so as to provide powerful information for policymakers and researchers. Each report focuses on a specific topic and is produced by a dedicated team of leading scholars on the theme in question. Teams are selected on the basis of an open call for tenders; there are two such calls a year. Tenders are subject to a thorough review process, as are the reports produced. (Full details are available on the IEA website.)

This seventh volume in the series is concerned with student motivation for learning mathematics. Student achievement in school depends on many school- and home-based factors, but also on individual motivation. Research into the motivational correlates of learning is not new, and it is clear that there is a positive relationship between motivation and learning. However, the links are generally found to be weak, and are imperfectly understood.

This study has taken the novel approach of focusing on students rather than motivation variables and, by identifying clusters of students with distinctive profiles in terms of these variables, sheds light on how motivation patterns relate differentially to achievement.

For analysis purposes the authors construe motivation in terms of three elements: enjoyment of, confidence in, and perceived value of mathematics (measures of each of these can be derived from the TIMSS background questionnaires). Cluster analysis across 12 education systems that took part at both grades four and eight in the TIMSS cycles of 1995, 2007, and 2015 identified a number of student clusters that remained stable across grades, over time, and across different systems. As might be expected, students scoring highly in all motivation elements also had high mathematics achievement levels. Critically, however, this study also delineates those clusters where the different elements of motivation went in contrary directions and provides a nuanced identification of the value of building student confidence in their mathematical abilities. Consistent patterns across grades, time, and systems show that confidence is a more important predictor of achievement than either enjoyment or perceived value.

Besides being a valuable addition to the literature on motivation and student learning, this report highlights the central importance of reinforcing students' competence in mathematics as a central construct toward building confidence. While it helps if students enjoy mathematics and appreciate its significance, in the absence of a well-founded confidence in mathematics, these factors have a weaker association with achievement.

Future publications in this series will include an in-depth investigation into the nature and extent of students' misconceptions and misunderstandings related to core concepts in mathematics and physics across grades four, eight, and 12, and a comprehensive analysis of gender differences in grade eight students' use and understanding of computer technologies.

Seamus Hegarty
Leslie Rutkowski
Series editors

Acknowledgments

This research was supported by a grant from the International Association for the Evaluation of Educational Achievement. We also thank Militsa Ivanova and Anastasios Markitsis for research assistance on this project.

Contents

1	Introduction to Motivational Profiles in TIMSS Mathematics	1
1.1	Motivation in Mathematics in Studies of Educational Achievement	1
1.2	A Person-Centered Approach to the Study of Motivation in TIMSS Mathematics	4
1.3	Potential to Expand the Current State of Research	5
1.4	Overview of This Book	5
	References	6
2	The Relationship of Motivation with Achievement in Mathematics	9
2.1	Introduction to Student Motivation	9
2.2	Theoretical Approaches to the Study of Motivation	11
2.2.1	Self-determination Theory	11
2.2.2	Expectancy-Value Theory	12
2.2.3	Self-efficacy Theory	13
2.2.4	Self-concept	13
2.2.5	Achievement Goal Theory	14
2.3	Measures of Motivation in TIMSS	14
2.4	The Relationship Between Motivation and Achievement	15
2.5	Self-reported Ratings of Motivation Across Education Systems	16
2.6	Self-reported Ratings of Motivation Across Ages	17
2.7	Another Approach to Studying the Motivation–Achievement Relationship	18
	References	19

3	Methodology: Cluster Analysis of Motivation Variables in the TIMSS Data	25
3.1	TIMSS Sampling	25
3.2	Jurisdictions Included in This Study	26
3.3	Instrumentation	28
3.3.1	Motivation Measures in the TIMSS 2015 Administration	29
3.3.2	Motivation Measures in the TIMSS 2007 Administration	29
3.3.3	Motivation Measures in the 1995 Administration	32
3.4	Other Variables Included in the Study	34
3.4.1	TIMSS Achievement Score Estimation	34
3.4.2	Other Variables of Interest	35
3.5	Analysis Technique	36
	References	39
4	Cluster Analysis Results for TIMSS 2015 Mathematics Motivation by Grade and Jurisdiction	41
4.1	Introducing the Person-Centered Approach	41
4.2	Cluster Analysis Results for the TIMSS 2015 Administration at Grade Four by Jurisdiction	42
4.2.1	Australia	42
4.2.2	Canada-Ontario	43
4.2.3	Canada-Quebec	45
4.2.4	England	46
4.2.5	Hong Kong	46
4.2.6	Hungary	48
4.2.7	Iran	49
4.2.8	Japan	50
4.2.9	Norway	51
4.2.10	Singapore	52
4.2.11	Slovenia	52
4.2.12	USA	54
4.3	Cluster Analysis Results for the TIMSS 2015 Administration at Grade Eight by Jurisdiction	55
4.3.1	Australia	56
4.3.2	Canada-Ontario	57
4.3.3	Canada-Quebec	59
4.3.4	England	59
4.3.5	Hong Kong	61
4.3.6	Hungary	61
4.3.7	Iran	63
4.3.8	Japan	65

4.3.9	Norway	66
4.3.10	Singapore	67
4.3.11	Slovenia	69
4.3.12	USA	70
5	Cluster Analysis Findings Over 20 Years of TIMSS	73
5.1	Summary of Cluster Analysis Results for Grade Four Samples	73
5.1.1	The TIMSS 1995 Administration	74
5.1.2	The TIMSS 2007 Administration	75
5.1.3	The TIMSS 2015 Administration	75
5.2	Summary of Cluster Analysis Results for Grade Eight Samples	76
5.2.1	The TIMSS 1995 Administration	76
5.2.2	The TIMSS 2007 Administration	77
5.2.3	The TIMSS 2015 Administration	78
5.3	Twenty-Year Patterns in TIMSS by Country and Grade	79
6	Insights from Motivational Profiles in TIMSS Mathematics	85
6.1	Examining the Role of Motivation in Educational Achievement	85
6.2	Clusters of Students Using Motivation Variables: A Person-Centered Approach	87
6.3	Motivation Clusters and Achievement	88
6.4	Motivation Clusters, and Student and Family Characteristics	89
6.5	Methodological Concerns	92
6.6	Concluding Remarks	93
	References	94
	Appendix A: IBM SPSS Code for the Two-Step Cluster Analysis	97
	Appendix B: TIMSS 1995 and 2007 Boxplots by Cluster for Each Jurisdiction	99
	Appendix C: TIMSS 1995 and 2007 Descriptive Statistics by Cluster for Each Jurisdiction	125

Chapter 1

Introduction to Motivational Profiles in TIMSS Mathematics



Abstract The role of motivation in educational achievement has been the focus of considerable research interest. While most empirical studies of the relationship between motivation and achievement use a variable-centered approach, this investigation takes a person-centered approach to identify student motivational profiles and examine their association with students' mathematics performance in International Association for the Evaluation of Educational Achievement's (IEA) Trends in International Mathematics and Science Study (TIMSS) and with other sociodemographic variables over time and across countries. The research aims to describe profiles with consistent or mixed scores on TIMSS motivation variables, and establish the relationship between profiles with mathematics achievement and with other sociodemographic variables. The motivation profiles and their characteristics are examined for students in grades four and eight, across 12 participating jurisdictions and over three administrations from 1995 to 2015.

Keywords Mathematics achievement · Motivational profiles · Sociodemographic variables · Student performance · Trends in International Mathematics and Science Study (TIMSS)

1.1 Motivation in Mathematics in Studies of Educational Achievement

In the study of school achievement, sociodemographic background variables have a large impact on test scores (see, e.g., Hattie 2009). Beliefs, values, and opinions students themselves have towards school subjects and assessment are less influential, but important correlates of achievement. In international large-scale assessments (ILSAs) for example, student motivation, self-efficacy, and self-concept in academic subjects are consistent predictors of student performance across almost all participating countries (Marsh et al. 2006, 2013). These predictors are important, because it is possible to support, coach, or develop students' adaptive beliefs about the purpose of assessment (Brown 2011), interest in school subjects (Alexander 2003), confidence in their own abilities (Bandura 1977), goals (Ryan and Deci 2000) and motives (Eccles and Wigfield 2002) for learning, and so on. However, while each of

these individual psychological and affective factors have been studied as predictors of performance, they are less often studied as patterns within individuals. Is it true that only students who consistently and systematically score highly on all factors will achieve the best academic performance? Is it possible that there are combinations of factors that generally support good academic achievement, and certain factors are less critical? We here aim to examine whether there are patterns of motivational factors that can be associated with higher test performance, whether these vary systematically with sociodemographic variables, and whether those patterns are universal and stable across age, time, and location.

The International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS) has included self-report measures of student attitudes and motivation in mathematics and science in the student background questionnaires administered together with the achievement tests since 1995. Competence, interest in science and mathematics, and positive affect and value ascribed to the two subjects, have been operationalized as self-report items, and, more recently, as scales. These constructs matter, partly because they are within the control of families and students themselves (and possibly even schools), and also because they contribute to life-long learning and the development of adults who are more capable of coping with the demands of modern society.

Various theoretical frameworks have posited the link between motivation to learn and academic success (e.g., Bandura 1997; Deci and Ryan 1985; Wigfield and Eccles 2000). Confidence perceptions as indicators of self-concept are thought to relate to engagement with purposeful behavior and success in academic tasks, and increased self-confidence is more likely to lead to successful outcomes. Ascribing value to a task and its outcome is another factor linked to academic performance that includes intrinsic characteristics like enjoyment, interest, and self-perceived importance, as well as costs and perceptions of usefulness. Moreover, these affective and motivational attributes are considered to be important, not just as predictors of achievement but also as valued schooling outcomes.

In a meta-analysis of 288 studies, Hattie (2009) reported that attitudes toward mathematics and science correlated with student achievement in both subjects. While this relationship has been characterized as positive and strong (Mullis et al. 2012), empirical evidence suggests a less pronounced network of associations. For example, in their multinational analyses of data from TIMSS and the Organisation for Economic Cooperation and Development's Programme for International Student Assessment (OECD's PISA), Lee and Stankov (2018), and Marsh et al. (2006, 2013) found weak correlations between value and affect for the subject with achievement for both mathematics and science; the only moderate to strong relationship was that between self-concept in the subject and achievement.

Developmental changes and group differences in motivational and affective variables, particularly with respect to mathematics, have also been explored in the literature. For example, amongst all school subjects examined, intrinsic motivation in mathematics has been identified as exhibiting the greatest decline from ages nine to 17 (Gottfried et al. 2001); interest and competency perceptions in mathematics have also been found to decline as students transition from primary to secondary

education (Fredericks and Eccles 2002). Longitudinal investigations have provided further supporting evidence that decreasing student achievement in mathematics is related to decline in motivation (Gottfried et al. 2007).

Country-level comparisons of motivational constructs have revealed important mean differences: percentages of students reporting low enjoyment in mathematics were low in countries that performed below the average level in TIMSS, while countries with high average scores had larger percentages of students expressing discontent (Mullis et al. 2016). This was found both in the 1995 administration of TIMSS and 20 years later in 2015. Similar results were found with an indicator of self-competence in mathematics. Employing more complex modeling approaches, Marsh et al. (2013) noted that mean motivational scores from four Arab countries were higher than those from four Anglo-Saxon countries, even though the mean differences in mathematics achievement followed the opposite pattern. Differential self-concept frames of reference models and cultural value orientation were provided as tentative explanations of this paradox.

However, these analyses depend on a detailed examination of each variable's relationship to achievement. The intercorrelations of these motivation variables are usually moderate (see, e.g., Marsh et al. 2013), suggesting that students do not score consistently high, medium, or low across the board. Thus, mixtures of motivational attitudes may create different profiles. For example, it is possible students who value a subject, might also rate themselves as incompetent and dislike the subject, raising the question as to whether valuing the subject can compensate for low interest and self-competence. The relative standing of these motivational variables may not be consistent. Consequently, in addition to students who have or who report consistent ratings for all such variables, whether high, medium, or low, the expectation is that there are also students with inconsistent ratings. Empirical observation may reveal patterns in motivational variables that are particularly related to achievement, but this interesting and novel topic has not been addressed in the literature. From an individual differences framework, student profiles may also differ in terms of sociodemographic characteristics (such as sex, age, ethnicity, or socioeconomic status). It may be that students with particular characteristics tend to aggregate into certain patterns of affect, motivation, and self-awareness. Better understanding of the interplay of factors affecting student achievement may thus lead to targeted interventions. We were unable to find descriptions and comparisons of student motivational profiles for students undertaking low-stakes formal testing across time at the country level in the literature. Hence, the TIMSS background questionnaire and achievement data provide a unique opportunity to employ a person-centered approach in a low-stakes context.

1.2 A Person-Centered Approach to the Study of Motivation in TIMSS Mathematics

Correlational analyses are traditionally employed in ILSA programs to examine the relationships between motivational variables and achievement outcomes. These analyses are variable-centered, since they model simple or multivariate relationships across individuals and serve a nomothetic function in theory building: how the hypothesized constructs covary, usually assuming linearity in the relationships, and build up a network of associations of these psychological characteristics.

An alternative approach is employed in this study: namely, we examine whether meaningful profiles can be empirically extracted from student self-reports on motivational and affective variables (e.g., students may report low self-concept, but high extrinsic value for mathematics). This person-centered approach, in which groups of students are formed based on similar profiles, is a less common analytic strategy and is complementary to the variable-centered approach (Loken and Molenaar 2008; Marsh et al. 2009). The hypothesis is that the various predictors interact in ways that will reveal stronger and weaker associations with achievement, compared to the average relationships observed in variable-centered approaches. For example, students might have positive self-concept, but low interest in mathematics; the impact of this mixed profile is not understood. A second hypothesis is that the composition of the clusters will differ by background and demographic characteristics (e.g., is it true that students from low socioeconomic status groups are less motivated to do mathematics than those from high socioeconomic status groups?) To the degree that societies differ in their motivational emphases toward testing (e.g., formal testing is highly valued in East Asia, while low-stakes tests are not particularly valued in Nordic countries) and the degree to which there is change in emphasis on formal testing within a society (e.g., Australia has relatively recently introduced formal school and system evaluation through an annual national assessment program of literacy and numeracy; see <https://www.nap.edu.au/>), the profiles of students who report having adaptive or maladaptive profiles may be expected to differ across countries. Separate descriptive analyses of each cluster will define the profile of students comprising each cluster and how the contributing factors relate to student performance.

The wealth of data available in TIMSS enables us to explore some additional topics. Firstly, within a country, clusters of students at grade four and grade eight can be compared to establish whether there are differences in national motivational profiles from childhood to early adolescence. Secondly, we will be using data spanning 20 years, from 1995 to 2015, maximizing the possibility that differences between countries and changes within countries over time can be observed and potentially understood.

1.3 Potential to Expand the Current State of Research

Interest in the relationship between motivational variables and achievement is not new. However, here we propose an alternative person-centered approach to this domain of study, designed to identify meaningful clusters of students depending on their motivational, competency, and affective responses. Assuming that there will be categorically different clusters of students, we anticipate that the composition of clusters will vary according to background and demographic measures, and hence that distributions of achievement by cluster will differ. We hypothesize that such findings may allow for interventions to be customized to suit students' motivational profiles.

The results may provide additional evidence about cohort changes in motivational profiles derived from ILSA surveys. Cross-sectional differences in the motivational profiles of grade four versus grade eight students may support the hypothesis of developmental variations in motivation profiles across countries (namely that adolescents tend to be less optimistic than younger children about school achievement and subjects). It may be that certain types of motivational decline (such as decline in self-rated competence) may actually be associated with improved performance, since it is well-established that, with increasing competence, self-assessments become less optimistic (Brown and Harris 2013). Identifying changes in the cohorts over 20 years will provide new empirical data on generational differences in motivation profiles. We examine initial cross-cultural differences and their potential links to country-level information and policy contexts.

1.4 Overview of This Book

In this book, we aim to identify patterns of motivational factors in mathematics, and compare clusters of students with different motivational profiles on achievement and sociodemographic variables. Adopting a person-centered approach, data from multiple education systems, administrations, and both grades assessed by TIMSS are used to examine whether patterns are universal and stable across age, time, and jurisdiction. Chapter 2 provides a detailed review of the research literature on the relationships between motivation and affect and mathematics performance, and we present relevant theoretical frameworks, as well as evidence from both small- and large-scale studies, including TIMSS. In Chap. 3, we provide an overview of the TIMSS samples and the measures we selected to address our research questions. We briefly describe the TIMSS sampling framework, and provide more specific information on the datasets we selected, which include data for both grade four and grade eight, and for the years 1995, 2007, and 2015, for all jurisdictions that participated in all three of these TIMSS cycles. We define the analytic approach we used to identify and evaluate clusters of students with similar motivation profiles, and provide examples of the software code we used for the analysis. In Chap. 4, we

report motivation clusters for each jurisdiction and grade in detail for the 2015 cycle of TIMSS administration; the results for earlier cycles of TIMSS are provided as Appendices. After describing the composition of clusters in terms of motivational characteristics, we compare clusters using demographic and achievement variables. Chapter 5 summarizes findings from all administrations and grades; selected cluster characteristics highlight trends across time and jurisdictions. Finally, in Chap. 6, we discuss the potential implications of our findings and the novel analytic approach, outlining limitations and possible future research directions.

References

- Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. London, UK: Worth Publishers.
- Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: A review of the students' conceptions of assessment inventory. *Educational Psychology*, 31(6), 731–748.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental Psychology*, 38(4), 519–533.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13.
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *International Journal of Behavioral Development*, 31(4), 317–327.
- Hattie, J. A. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon, UK: Routledge.
- Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64.
- Loken, E., & Molenaar, P. (2008). Categories of continua? The correspondence between mixture models and factor models. In G. R. Hancock, & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 277–297). Charlotte, NC: Information Age.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., ... & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108–128.
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311–360.

- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, 16(2), 191–225.
- Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf.
- Mullis, I. V. S., Martin, M. O., & Loveless, T. (2016). *20 years of TIMSS: International trends in mathematics and science achievement, curriculum, and instruction*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timss2015.org/timss2015/wp-content/uploads/2016/T15-20-years-of-TIMSS.pdf>.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 2

The Relationship of Motivation with Achievement in Mathematics



Abstract Students’ motivations and attitudes influence their learning of and performance on assessments of school subjects, including mathematics. Relevant theories (e.g., self-determination, expectancy-value, self-efficacy, self-concept, and achievement goal theory) capture key motivational factors that are important to learning and achievement. The theories and related empirical studies that show how motivation factors relate to performance are comprehensively reviewed in this chapter. Generally, students who demonstrate greater self-efficacy and interest in mathematics, and those who value the subject more highly achieve better outcomes, although the relationships may often be modest. Correlational and experimental evidence for the association of motivation with achievement is presented with reference to longitudinal and cross-cultural comparisons, as well as to findings from international large-scale assessments, such as IEA’s Trends in International Mathematics and Science Study (TIMSS). Cluster analysis is one tool that may be used to understand the person-centered conditions that are most strongly associated with specific motivational factors related to TIMSS achievement scores.

Keywords Achievement goal theory · Expectancy-value theory · Person-centered approach · Self-concept · Self-determination theory · Self-efficacy · Student motivation theories

2.1 Introduction to Student Motivation

If there is a cornerstone in the science of human behavior, it must be the field of motivation. Motivational theories ask a fundamental question, namely: what moves a person. They are concerned with the prime force at work in human nature and human culture (Ryan 1998, p. 114).

The study of motivation is central to the science of human behavior in general, and Ryan’s (1998) statement is also relevant to educational science. By motivation, we mean the reasons, purposes, intentions, goals, feelings, intuitions, values, beliefs, and attitudes that humans use to explain why they do the things they do (Mercier and Sperber 2017). Simply put, “a person who feels no impetus or inspiration to act is thus characterized as unmotivated, whereas someone who is energized or activated toward

an end is considered motivated” (Ryan and Deci 2000, p. 54). While environmental and genetic factors may shape human behavior, individual motivation contributes to behavior and outcomes.

In educational settings, the study of student motivation is central, because this interacts with educational choices and achievement behavior in important ways (Bandura 1997; Deci and Ryan 1985; Hattie 2009; Wigfield and Eccles 2002). Students need to learn the skills and competencies expected of them and, to do so, they need the will to learn and perform well in educational assessment settings (Eklöf 2010). Learning science has gone further and pointed to the need for students to gain skills, the will to learn, and experience a motivating thrill in learning that enables them to persist when learning becomes difficult (Hattie and Donoghue 2016).

Motivation is not a static characteristic, but malleable, and differs between and within individuals across different domains, activities, and contexts (Ryan and Deci 2000). Some students learn to achieve extrinsic rewards, while others learn for more intrinsic reasons. The strength of these goal motivations can vary across time, according to the context, and by a student’s level of maturity, culture, and individual background demographic characteristics. Thus, motivation matters because, unlike societal habitus or genetic inheritance, individuals seem to be able to select from a range of options, as well as monitor the efficacy of motivation and exercise control over their motives.

Investigations into the mechanisms through which motivation and performance interact, and their causal precedence, remain an important area of research (see, e.g., Cerasoli and Ford 2014; Marsh and Craven 2006). Success or failure interacts with motivating beliefs such that certain motivations and actions are reinforced or inhibited in response to outcomes. For example, if self-efficacy judgments are unrealistic because the self-assessment of competence was not accurate (which occurs especially among novices; Brown and Harris 2013), it is likely that such students will lose motivation or reduce the effort required to undertake a task. However, if the self-efficacy rating accurately predicted actual performance, this could reinforce the students’ sense of knowing and could lead to appropriate decisions about continued effort in specific learning domains. Consequently, with multiple aligned or misaligned interactions between motivational beliefs (e.g., self-efficacy, interest, or value) and actual performance, students become reinforced in their motivational stance. Of course, the causal ordering in such reciprocal relations is not well-established, as it is difficult to ascertain a true zero-point baseline. Even if the causal ordering of constructs is not clearly established, there is ample evidence that there are links between knowledge, abilities, academic behaviors, and performance on the one hand, and aspects of motivation on the other (Hattie 2009; Robbins et al. 2004).

The exploration of student motivation in schooling contexts has attracted a great deal of research, and motivational constructs have been evaluated in the TIMSS studies of achievement in mathematics and science (Hooper et al. 2017). Despite the large body of research on achievement motivation, there is still much to be learned, not least when it comes to motivational patterns within individuals, across culture, time, and age groups, and how such motivational patterns inform differences

in achievement distributions across nations and changes in these distributions over time. Identifying information that is potentially hidden in the correlational approach that is usually undertaken in secondary analyses of ILSA data is important; thus, we will begin by briefly describing major frameworks on motivation before examining their relevance to the motivation constructs included in TIMSS questionnaires.

2.2 Theoretical Approaches to the Study of Motivation

Motivation, as a latent construct, is not directly observable, thus theoretical framing matters to its measurement. Motivational theories are concerned with understanding what makes people act in motivated ways, and what makes an individual choose the direction and intensity of actions. What then are the more influential motivational theories, and how are they reflected in the measures of motivation used in TIMSS?

With the advent of modern psychology in the late 19th century, interest developed in the possible factors involved in human motivation. Since then, a “myriad” (Gottfried 2009, p. 464) of theories and explanatory frameworks for motivation have evolved. These have stressed inner needs, drives, instincts, and relationships between stimuli, response, and reinforcement as the initiators of motivated behavior. The modern motivational paradigm is dominated by cognitive theories, which claim that individuals’ thoughts, beliefs, and emotions together influence motivation (see Schunk et al. 2010; Wigfield and Eccles 2002; Wentzel and Wigfield 2009). While older theories of motivation (Atkinson 1957) often saw motivation as emanating from needs, or a “drive,” contemporary theories tend to stress goals as important motivators; this distinction between push (drive) and pull (goal) is significant. Five key theories are relevant in the context of the measurement of motivation in TIMSS.

2.2.1 *Self-determination Theory*

One of the more influential motivational theories is self-determination theory (SDT). In this (macro-)theory, a distinction is made between different types of motivation based on the different kinds of reasons or goals that give rise to motivated action. SDT is concerned with people’s inherent growth tendencies and innate psychological needs (i.e., competence, autonomy, and psychological relatedness) that evolve in interaction with the surrounding social context. SDT focuses on the degree to which an individual’s behavior is self-motivated and self-determined. Although rather complex in detail (see Ryan and Deci 2002, 2017), a basic distinction is often made between extrinsic and intrinsic motivation (Ryan and Deci 2000).

Extrinsic motivation refers to goals or reasons related to reaching an external reward (e.g., money, praise from others, good grades) or avoiding negative consequences (e.g., embarrassment upon getting a poor test result, or not being allowed to visit a friend unless homework has been completed as expected).

According to Deci and Ryan (1985), extrinsic motivation can become more or less internalized and is not, as often described, antithetical to intrinsic motivation. In other words, it can be intrinsic to the self to aspire to external rewards. Intrinsic motivation, in contrast, is when an action is performed “for its own sake,” without expectation of external rewards. It can concern things such as personal development, enjoyment of exploring that leads to feelings of “internal rewards,” reaching personal goals, mental satisfaction, physical well-being, or enjoyment in the moment. The focus can be on the process more than the result, though the result itself may be invisible to external parties. Intrinsic and internalized (self-determined) motivation is often linked to stronger devotion, greater persistence, and deeper processing. In practice, motivated behavior is often guided by both internal and external factors. However, individuals will differ in the balance between internal and external motivations, and the degree of internalization of extrinsically-focused motivating factors may appear to differ even within social contexts.

TIMSS has measured attitudes and motivation towards learning and achieving in mathematics and science from 1995 and onwards (Hooper et al. 2017). However, it is not clear if and how operationalization of SDT into TIMSS survey items has been guided by theory. In the recent assessment frameworks, there was reference to SDT when motivational constructs were described (Hooper et al. 2013, 2017). Enjoyment, confidence, and value for mathematics and science are constructs operationalized and measured in the TIMSS background questionnaires. Inspection of the scales indicates that enjoyment items relate to intrinsic motivation and some value items relate to extrinsic motivation.

2.2.2 Expectancy-Value Theory

Previous research has shown that it is possible to interpret the motivational scales used in TIMSS within an expectancy-value framework (Eklöf 2007). The expectancy-value theory (EVT) of achievement motivation originates from the first half of the 20th century (Atkinson 1957; Weiner 1992), but the most widely used expectancy-value model derives from the more recent work of Eccles and Wigfield (Eccles and Wigfield 2002; Wigfield 1994; Wigfield and Eccles 2002). This model focuses on the role of students’ expectancies for academic success and their perceived value for academic tasks; it is based on personality, social, and developmental psychology (Pintrich and Schunk 2002). EVT has two core components: an expectancy component that corresponds to the question “Can I do this task?” (namely, student perceptions of their abilities), and a value component that corresponds to the question “Do I want to do this task and why?” The expectancy component in the model refers to the individual’s beliefs and judgments about his or her capabilities to do a task and succeed at it, with obvious correspondence with constructs from other theoretical traditions such as self-schemata, self-concept, or self-efficacy. The value component in the model refers to the various reasons individuals have for engaging in a task or not, and the strength of those values. Both of these components have been shown to be important predictors of achievement behavior (Wigfield and Eccles 1992).

The different value components in the model are attainment value (perceived importance of doing well on a task), intrinsic value (interest/enjoyment in doing a task), utility value (usefulness in terms of the individual's future goals), and cost (perceived amount of effort required for a task). Intrinsic value is conceptually similar to intrinsic interest in SDT (Deci and Ryan 1985), while utility value resembles the extrinsic motivation component in SDT.

Expectancy beliefs, including self-concept, ability perceptions, and expectancy for success, predict actual achievement in terms of grades and performance on standardized tests. Values have been shown to correlate positively with actual achievement, but when both expectancy beliefs and values are used simultaneously to predict achievement, expectancy beliefs remain significant predictors, while values become not significant predictors. However, in terms of intentions to take future courses and actual enrollment in those courses, value beliefs are actually better predictors than expectancy beliefs (Eccles 1983; Meece et al. 1990; Wigfield and Eccles 1992). Confidence and enjoyment for mathematics and science have been measured in TIMSS questionnaires using several items in the past, although more recently, they have been measured as separate, multi-item scales; grade eight students are asked to report how much they value the two subjects.

2.2.3 Self-efficacy Theory

Self-efficacy can be understood as the individual's conviction that he or she is capable of successfully performing a given task (Bandura 1997). This is analogous to "Can I do this task?" in the terminology of EVT. Self-efficacy theory is grounded in a larger theoretical framework known as social cognitive theory, which assumes that human achievement depends on interactions between an individual's behaviors, personal factors (e.g., thoughts and behaviors), and environmental conditions. TIMSS does not measure self-efficacy, although we found references to Bandura's work in the theoretical frameworks (e.g., see Hooper et al. 2013).

2.2.4 Self-concept

Related to self-efficacy, and not always easy to separate empirically, is the theoretically different and broader construct of self-concept. Self-concept can be defined as an individual's "collective self-perceptions that are formed through experiences with and interpretations of the environment" (Wigfield and Eccles 2002, p. 16). Self-concept then is how individuals perceive and evaluate themselves: questions such as "Who am I?" and "What kind of person am I?" are the foci of self-concept (Hattie 1992). Self-concept is heavily influenced by reinforcements and evaluations by significant others (Shavelson and Bolus 1982) and it is multi-dimensional in the sense that self-concept differs according to context; an individual's self-concept could differ for mathematics, reading, or academics in general

(Marsh 1990). Self-concept differs from self-efficacy both in specificity and content (Michaelides 2008); the former focuses on “Who I am,” while the latter focuses on “How sure am I that I can do something?” Aspects of self-concept and self-efficacy are also incorporated as parts of other models of motivation, such as EVT. In recent cycles of TIMSS, the measure of student confidence in mathematics and science included items that could be related to aspects of self-concept.

2.2.5 Achievement Goal Theory

Achievement goal theory (Ames 1992; Dweck and Leggett 1988; Elliott and Dweck 1988; Pintrich 2000) assumes that students have different reasons for engaging or not engaging in learning and school work. These reasons affect what, how, and why students learn (Patrick et al. 2011) and how they subsequently perform. Typically, two different sets of achievement goals are identified: (1) task and ego goals (Nicholls 1984) or (2) mastery and performance goals (Dweck and Leggett 1988). The primary goal of a mastery-oriented person is the learning and mastery of the task for its own sake (similar to intrinsic motivation in SDT and intrinsic value in EVT). Mastery goals reflect a desire to achieve competence in terms of set criteria or task mastery. They rely on comparisons with the explicit requirements of the task and/or internal comparisons with an individual’s past or potential attainment. Performance goals, in contrast, reflect a desire to achieve competence to a degree that is relative to the performance of others. The goal here is to do well and gain the rewards associated with high performance. The desire to attain high performance sometimes leads to strategic behaviors that can involve making learning more difficult for competitors (Murdock et al. 2016). These two goal orientations are associated with different consequences in achievement context, with mastery being associated with higher performance than performance orientation.

2.3 Measures of Motivation in TIMSS

TIMSS administers questionnaires to different actors in the school system, including students, school leaders, teachers, and parents. Over the years, the questionnaires have changed, as have the components of motivation that are assessed (see Mullis et al. 2017). As our study aimed to evaluate trends in motivation, we restricted ourselves to measures of motivation that were present in all the TIMSS cycles included in the study. Therefore, we focused on the constructs “Students like learning mathematics,” “Students value mathematics,” and “Students are confident in mathematics,” which have all been operationalized in recent administrations. In the TIMSS frameworks (see, e.g., Hooper et al. 2013, 2017), the construct “Students like learning mathematics” is interpreted as a measure encompassing intrinsic motivation, while “Students value mathematics” is interpreted as a measure comprising items

related to extrinsic motivation, attainment, and utility value. “Students’ confidence in mathematics” is intended as a measure of student self-concept. The deeper theoretical rationale behind these interpretations is not made clear in the frameworks. As described in Sect. 2.2, the frameworks reference SDT, intrinsic and extrinsic motivation, and the work of Deci and Ryan (1985). Self-concept is also briefly mentioned, with reference to Bandura (1997) and the likelihood of successfully completing the school task, as well as to the multi-dimensionality and subject-specificity of self-concept and the work of Marsh et al. (2000, 2006, 2008, 2013).

The empirical rationale is made clear through evidence that these constructs have shown positive relationships with performance in previous TIMSS studies. Most of the theories we have described could be used as interpretative frameworks for the motivational measures in TIMSS. As we mentioned, motivation is a latent construct and, thus, cannot be observed directly. It follows that constructs such as self-concept and task value do not equal motivation, but are constructs that have been shown to be predictive of achievement motivation (Wigfield and Eccles 2002), which in turn has been shown to be predictive of achievement choices and performance.

The relationship between student motivation and their achievement has been the topic of many reports. In reviewing the literature, we focused on studies performed in the TIMSS context, but also identified other relevant studies.

2.4 The Relationship Between Motivation and Achievement

The explanation and prediction of achievement is important to most educational motivation theories (Meece et al. 2006). A vast number of studies have explored components of student motivation and their relationships with performance, and meta-analyses have shown a positive (but often modest) relationship between student motivation (measured in different ways, using different theoretical frameworks) and scholastic achievement (Hattie 2009; Karadag 2017; Lee and Stankov 2018; Osborne et al. 2003). Throughout most of the theoretical and empirical work, the main message is that adaptive and high levels of motivation positively correlate with better learning and achievement. We do not aim to comprehensively review the motivation research field, instead we focus on studies exploring the variables that are assessed in TIMSS and other ILSAs, studies looking at patterns of motivation as children grow older, and a number of studies using person-centered approaches.

Research within the EVT framework suggests that mathematics and science achievement are positively associated with young peoples’ values and self-concepts (Casey et al. 1997; Frome and Eccles 1998; Wigfield et al. 1991). In the TIMSS context, the self-concept measure has received particular attention from researchers. A positive self-concept has been shown to be positively related to achievement in TIMSS (Eklöf 2008; Shen and Pedulla 2000), and “results from six previous TIMSS cycles have shown a strong relationship between students’ academic self-concepts and their achievement” (Hooper et al. 2017, p. 72). The relation between

perceived task value and performance is often less pronounced, but in the TIMSS documentation, it is stated that “TIMSS results have consistently shown a strong relationship between students valuing the subject and their achievement” (Hooper et al. 2017, p. 72; see also Mullis et al. 2004). Similar findings have been reported for TIMSS science (Papanastasiou and Zembylas 2004) and TIMSS mathematics (Patterson et al. 2003). Conversely, researchers have reported only weak relationships between the valuing mathematics scale and mathematics score for some education systems participating in TIMSS (Eklöf 2008; Grønmo et al. 2004; Thomson and Fleming 2004), in particular when other variables have been accounted for in regression models. Multinational analyses from TIMSS and PISA have found rather weak relationships between valuing the subject and achievement (Lee and Stankov 2018; Marsh et al. 2006, 2013).

2.5 Self-reported Ratings of Motivation Across Education Systems

Despite some mixed evidence regarding the strength of the relationship between student motivation and student achievement within education systems participating in ILSAs, the relationship between the self-report measures of motivation and student performance is usually positive. However, at the between-country level these relationships are often non-significant or even negative (Artelt 2005; Shen and Pedulla 2000). This seemingly paradoxical finding has been consistently reported in analyses of assessment data. For example, students in poorly-performing education systems often report a positive self-concept in school subjects, while students in the highest-performing education systems often report a low self-concept. Herbert Marsh proposed a “big-fish-little-pond-effect” (or internal versus external frame of reference) to understand this phenomenon (Marsh 1987; Marsh et al. 2000, 2008).

A six-nation study using TIMSS 1995 data showed large variation between education systems in terms of reports of liking mathematics and self-concept (Boe et al. 1999). Within countries, motivational factors had differing effects on performance but, in all countries, self-concept was a strong predictor of performance. Liking mathematics and science was also positively related to performance, but, when self-concept was added to the model, the effect of liking the subject disappeared in some education systems. It was concluded that liking mathematics is more likely to be a function of a positive mathematics self-concept than the reverse. Even if modest in their effects on performance, highly positive attitudes towards mathematics and science can make a substantial change to students’ learning and achievement (Boe et al. 1999).

Differing motivational ratings between students from different cultures and countries can also be due to differences in response patterns. There appear to be cultural differences in how students treat the response scales (e.g., extreme or midpoint responding) and those seem to relate to cultural values and affluence (He et al. 2014; Shulruf et al. 2011). For example, within New Zealand’s national

Assessment Tools for Teaching and Learning test norming of reading, writing, and mathematics, the relationship of liking and self-efficacy to achievement was statistically non-significant for two minority ethnic groups (i.e., the Pasifika and Māori) while it was positive for majority Pākehā and minority Asian groups ('Otunuku and Brown 2007). Survey items are vulnerable to adaptations, and, even if appropriate, terms and labels might have different meanings to different people. Studies using TIMSS 2011 data have shown that negatively worded attitude items were associated with response bias that differed across countries (Michaelides 2019; Tee and Subramaniam 2018), a result that is consistently reported in the wider literature (see, e.g., Brown 2004).

2.6 Self-reported Ratings of Motivation Across Ages

Developmental changes in motivation in terms of expectancies and values have been rather extensively studied by Eccles and Wigfield and their colleagues (Eccles et al. 1997; Simpkins et al. 2006; Wigfield et al. 1991). These studies indicate that the mean level of self-perceptions of ability decreases as children move into adolescence. At the same time, the accuracy of self-perceptions seems to improve with age, partly because students base their self-perceptions on their interaction with their environment according to this social-cognitive perspective. Having fewer instances to relate their ability to objective criteria, younger student self-perceptions are less accurate; a widely reported phenomenon in studies of student self-assessment (Brown and Harris 2013). Perceptions of task value seem to follow the same developmental pattern. Thus, research has shown that younger children are more positive in terms of self-concept, liking, and valuing the subject than older students. This seems to be true also in the TIMSS context (Eklöf and Costa 2018).

Intrinsic motivation in mathematics has been found to have the greatest decline among school subjects from childhood to adolescence (Gottfried et al. 2001); interest and competency perceptions in mathematics have also been found to decline as students transition from primary to secondary education (Fredericks and Eccles 2002). Indeed, in a New Zealand cross-sectional study of student attitudes towards mathematics, the influence of liking a subject to achievement in that subject shifted from statistically non-significant among elementary school students to becoming negative among high school students, whereas self-efficacy was consistently a positive predictor in both samples (Winheller et al. 2013). Longitudinal investigations have supported the idea that achievement is related to the decline of mathematics motivation (Gottfried et al. 2007). As students gain more experience with a subject and realize their strengths and weaknesses relative to objective performance information, their interest and enjoyment on the whole is expected to decline. For example, in a small, two-year longitudinal study of American high school student attitudes and cheating behaviors, performance goals and attainment values declined by trivial amounts ($|d| < 0.10$) and attainment values had a weaker influence on grade

point average at the second time point ($\beta = 0.18$) than at the beginning of the study ($\beta = 0.35$) (Stephens and Brown 2013), demonstrating autoregressive characteristics of the motivational belief upon achievement.

2.7 Another Approach to Studying the Motivation–Achievement Relationship

Variable-centered approaches drawing on the linear model and using correlation and regression techniques are typically encountered in studies on the relationships between motivational variables and educational outcomes. A less common and complementary approach is to search for meaningful groups of individuals within a sample, groups that share similar profiles. We opted to use a person-centered approach (Loken and Molenaar 2008) as an alternative way of uncovering motivational and affective student profiles. The individual student constitutes the unit of analysis in this mode of research.

A number of studies have investigated motivational profiles of students in other contexts using a person-oriented approach, and valuable information can be revealed by this less traditional method. Motivational profiles can be homogeneous (i.e., consistently high, medium, or low on all dimensions) or heterogeneous. For instance, in the context of second language learning, there were homogeneous as well as heterogeneous clusters, that were characterized, for example, by students who had a positive attitude toward learning a second language but did not find it relevant to their future professional life (Csizér and Dörnyei 2005). A cluster analysis of achievement goal orientation variables with elementary school students (Meece and Holt 1993) revealed distinct groups with higher achievement outcomes when mastery goals were highly endorsed compared to other groups with different levels of goal orientations. Ratelle et al. (2007) reported distinct groups of high-school students with varying levels of autonomous, controlled, and amotivated regulation, which differed in their academic achievement. They also reported different groupings with a college sample, indicating that the context and/or developmental factors seem to matter.

The volatility of motivational profiles was reported even within the duration of a single course using a sample of engineering students (Dillon and Stolk 2012), and Bråten and Olaussen (2005) noted changing levels of motivation within an academic year in clusters of nursing and business administration students. In contrast, student clusters based on achievement goal orientation measures were found to be relatively stable across time in lower- and upper-school students (Tuominen-Soini et al. 2011). Baker and Wigfield (1999) applied cluster analysis to multiple reading motivation variables and extracted seven distinct clusters of students that differed in their reading activity but were less variable in their reading achievement. They concluded that, rather than thinking of children as either high or low in motivation, it was important to understand that many children have a mixture of motivational characteristics. Thus, the empirical literature seems to support the view that motivation is state-like, in that it is dynamic in response to context and is not necessarily homogenous within people across time or context.

Cluster analysis has been applied to the PISA data to explore student profiles of interests in science issues (Olsen and Lie 2011), and school engagement in Finland (Linnakylä and Malin 2008). In our review of the literature, we were unable to identify any other studies related to motivation that used a person-centered approach to analyze ILSA data; applying this method to the TIMSS data should thus provide novel insights.

References

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- Artelt, C. (2005). Cross-cultural approaches to measuring motivation. *Educational Assessment*, 10(3), 231–255.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behaviour. *Psychological Review*, 64(6), 359–372.
- Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, 34(4), 452–477.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman and Company.
- Boe, E. E., Turner, H. M., May, H., Leow, C., & Barkanic, G. (1999). The role of student attitudes and beliefs about mathematics and science learning in academic achievement: Evidence from TIMSS for six nations. *CRESPE Data Analysis Report*. Nashville, TN: Consortium for Risk Evaluation with Stakeholder Participation. Retrieved from http://repository.upenn.edu/gse_pubs/413.
- Bråten, I., & Olaussen, B. S. (2005). Profiling individual differences in student motivation: A longitudinal cluster-analytic study in different academic contexts. *Contemporary Educational Psychology*, 30(3), 359–396.
- Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, 94(3), 1015–1024.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, 33(4), 669–680.
- Cerasoli, C. P., & Ford, M. T. (2014). Intrinsic motivation, performance, and the mediating role of mastery goal orientation: A test of self-determination theory. *Journal of Psychology*, 148(3), 267–286.
- Csizér, K., & Dörnyei, Z. (2005). Language learners' motivational profiles and their motivated learning behavior. *Language Learning*, 55(4), 613–659.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Dillon, A., & Stolk, J. (2012). The students are unstable! Cluster analysis of motivation and early implications for educational research and practice. *2012 Frontiers in Education Conference Proceedings*. Piscataway, NJ: IEEE. Retrieved from <https://ieeexplore.ieee.org/document/6462485/authors>.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273.
- Eccles, J. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: Freeman.

- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Eccles, J. S., Lord, S. E., Roeser, R. W., Barber, B. L., & Jozefowicz, D. (1997). The association of school transitions in early adolescence with developmental trajectories through high school. In J. Schulenberg, J. L. Maggs, & K. Hurrelmann (Eds.), *Health risks and developmental transitions during adolescence* (pp. 283–320). New York, NY: Cambridge University Press.
- Eklöf, H. (2007). Self-concept and valuing of mathematics in TIMSS 2003: Scale structure and relation to performance in a Swedish setting. *Scandinavian Journal of Educational Research*, 51(3), 297–313.
- Eklöf, H. (2008). Test-taking motivation on low-stakes tests: A Swedish TIMSS example. In M. von Davier, & D. Hastedt (Eds.), *Issues and methodologies in large-scale assessments, IERI Monograph Series, Vol. 1*. Hamburg, Germany: IEA-ETS Research Institute. Retrieved from <http://www.ierinstitute.org/dissemination-area.html>.
- Eklöf, H. (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy and Practice*, 17(4), 345–356.
- Eklöf, H., & Costa, D. R. (2018). *Test-taking motivation in Swedish TIMSS and TIMSS Advanced: Findings across cohorts and over time*. Paper presented at the 2018 Annual Meeting of the National Council for Measurement in Education, New York, NY, April 12–16.
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5–12.
- Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental Psychology*, 38(4), 519–533.
- Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, 74(2), 435–452.
- Gottfried, A. E. (2009). Commentary: The role of environment in contextual and social influences on motivation: Generalities, specificities, and causality. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 463–476). New York, NY: Routledge.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13.
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *International Journal of Behavioral Development*, 31(4), 317–327.
- Grønmo, L. S., Bergern, O. K., Kjaernsli, M., Lie, S., & Turmo, A. (2004). *Hva i all Verden har skjedd i realfagena: Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2003* [What on earth has happened in the mathematics and science subjects: Norwegian students' achievement in mathematics and science in TIMSS 2003; in Norwegian]. Oslo, Norway: Oslo University, Department of Teacher Education and School Development.
- Hattie, J. A. C. (1992). *Self-concept*. Hillsdale, NJ: LEA.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
- Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: a synthesis and conceptual model. *npj Science Of Learning*, 1, 16013. Retrieved from <http://www.nature.com/articles/npjscilearn201613>.
- He, J., Dominguez Espinosa, A., Poortinga, Y. H., & Van de Vijver, F. J. R. (2014). Acquiescent and socially desirable response styles in cross-cultural value surveys. In L. T. B. Jackson, D. Meiring, F. J. R. Van de Vijver, E. Idemudia, & W. K. Gabrenya Jr (Eds.), *Toward sustainable development through nurturing diversity* (pp. 98–111). Melbourne, FL: International Association for Cross-Cultural Psychology. Retrieved from http://www.iaccp.org/sites/default/files/stellenbosch_pdf/He.pdf.

- Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 context questionnaire framework. In I. V. S. Mullis, & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 61–82). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Hooper, M., Mullis, I. V. S., Martin, M. O., & Fishbein, B. (2017). TIMSS 2019 context questionnaire framework. In I. V. S. Mullis, & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 57–78). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: <http://timssandpirls.bc.edu/timss2019/frameworks/>.
- Karadag, E. (Ed.). (2017). *The factors effecting student achievement: Meta-analysis of empirical studies*. Cham, Switzerland: Springer International Publishing. Retrieved from <https://www.springer.com/gp/book/9783319560823>.
- Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64.
- Linnakylä, P., & Malin, A. (2008). Finnish students' school engagement profiles in the light of PISA 2003. *Scandinavian Journal of Educational Research*, 52(6), 583–602.
- Loken, E., & Molenaar, P. (2008). Categories of continua? The correspondence between mixture models and factor models. In G. R. Hancock, & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 277–297). Charlotte, NC: Information Age.
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295.
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82(4), 623–636.
- Marsh, H. W., Abduljbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C.,... & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108–128.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163.
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311–360.
- Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, 78(2), 337–349.
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20(3), 319–350.
- Meece, J. L., & Holt, K. (1993). A pattern analysis of students' achievement goals. *Journal of Educational Psychology*, 85(4), 582–590.
- Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology*, 57, 487–503.
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, 82(1), 68–81.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Michaelides, M. P. (2008). Emerging themes from early research on self-efficacy beliefs in school mathematics. *Electronic Journal of Research in Educational Psychology*, 6(1), 219–234.
- Michaelides, M. P. (2019). Negative keying effects in the factor structure of TIMSS 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education* (in press).
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

- Mullis, I. V. S., Martin, M. O., & Hooper, M. (2017). Measuring changing educational contexts in a changing world: Evolution of the TIMSS and PIRLS questionnaires. In M. Rosén, K. Y. Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes: A festschrift in honour of Jan-Eric Gustafsson* (pp. 207–222). Cham, Switzerland: Springer International Publishing.
- Murdock, T. B., Stephens, J. M., & Grotewiel, M. M. (2016). Student dishonesty in the face of assessment: Who, why, and what we can do about it. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 186–203). New York, NY: Routledge.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*(3), 328–346.
- Olsen, R. V., & Lie, S. (2011). Profiles of students' interest in science issues around the world: Analysis of data from PISA 2006. *International Journal of Science Education*, *33*(1), 97–120.
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.
- Otunuku, M., & Brown, G. T. L. (2007). Tongan students' attitudes towards their subjects in New Zealand relative to their academic achievement. *Asia Pacific Education Review*, *8*(1), 117–128.
- Papanastasiou, E. C., & Zembylas, M. (2004). Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. *International Journal of Science Education*, *26*(3), 259–280.
- Patrick, H., Kaplan, A., & Ryan, A. M. (2011). Positive classroom motivational environments: Convergence between mastery goal structure and classroom social climate. *Journal of Educational Psychology*, *103*(2), 367–382.
- Patterson, M., Perry, E., Decker, C., Eckert, R., Klaus, S., Wendling, L., & Papanastasiou, E. C. (2003). Factors associated with high school mathematics performance in the United States. *Studies in Educational Evaluation*, *29*(2), 91–108.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Ziedner (Eds.), *Handbook of self-regulation* (pp. 451–502). London, UK: Academic Press.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (Second ed.). New Jersey, NJ: Merrill Prentice Hall.
- Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. *Journal of Educational Psychology*, *99*(4), 734–746.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288.
- Ryan, R. M. (1998). Commentary: Human psychological needs and the issues of volition, control, and outcome focus. In J. Heckhausen, & C. S. Dweck (Eds.), *Motivation and self-regulation across the life span* (pp. 114–133). New York, NY: Cambridge University Press.
- Ryan, R. M., & Deci, E. L. (Eds.). (2002). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivation: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York, NY: Guilford Press.
- Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2010). *Motivation in education: Theory, research, and applications* (Third ed.). London, UK: Pearson Education.
- Shavelson, R. J., & Bolus, R. (1982). Self concept: The interplay of theory and methods. *Journal of Educational Psychology*, *74*(1), 3–17.
- Shen, C., & Pedulla, J. J. (2000). The relationship between students' achievement and their self-perceptions of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education: Principles, Policy and Practice*, *7*(2), 237–253.

- Shulruf, B., Hattie, J., & Dixon, R. (2011). Intertwinement of individualist and collectivist attributes and response sets. *Journal of Social, Evolutionary, and Cultural Psychology*, 5(1), 51–65.
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1), 70–83.
- Stephens, J. M., & Brown, G. T. L. (2013). *Continuity and change: A longitudinal study of student motivation, misconduct and achievement during high school*. Paper presented at the Social Psychology of the Classroom International Conference, Auckland, New Zealand, July 15–18.
- Tee, O. P., & Subramaniam, R. (2018). Comparative study of middle school students' attitudes towards science: Rasch analysis of entire TIMSS 2011 attitudinal data for England, Singapore and the USA as well as psychometric properties of attitudes scales. *International Journal of Science Education*, 40(3), 268–290.
- Thomson, S., & Fleming, N. (2004). *Summing it up: Mathematics achievement in Australian schools in TIMSS 2003*. TIMSS Australia monograph no. 6. Camberwell, Australia: Australian Council for Educational Research.
- Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2011). Stability and change in achievement goal orientations: A person-centered approach. *Contemporary Educational Psychology*, 36(2), 82–100.
- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Thousand Oaks, CA: Sage.
- Wentzel, K. R., & Wigfield, A. (2009). *Handbook of motivation in school*. New York, NY: Routledge.
- Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6, 49–78.
- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12(3), 265–310.
- Wigfield, A., & Eccles, J. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield, & J. Eccles (Eds.), *Development of achievement motivation* (pp. 92–120). New York, NY: Academic Press.
- Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27(4), 552–565.
- Winheller, S., Hattie, J., & Brown, G. T. L. (2013). Factors influencing early adolescents' mathematics achievement: High-quality teaching rather than relationships. *Learning Environments Research*, 16(1), 49–69.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

Methodology: Cluster Analysis of Motivation Variables in the TIMSS Data



Abstract This chapter begins with a description of the IEA’s Trends in International Mathematics and Science Study (TIMSS) sampling framework. The research study was based on data from three cycles of TIMSS collected at grades four and eight from 12 jurisdictions (Australia, England, Hong Kong, Hungary, Iran, Japan, Norway, Ontario, Quebec, Singapore, Slovenia, and the USA) that participated at both grades in 1995, 2007, and 2015. The motivation variables available in each cycle of administration are outlined, together with how conceptually similar measures for value, enjoyment, and self-confidence in mathematics were constructed; descriptions of demographic and achievement measures used in the analysis are also provided. Two-step cluster analysis was used to create separate profiles of student motivation for each set of data. The characteristics of each motivational cluster were evaluated to ascertain whether differences in cluster membership were related to student background variables (such as sex, time on homework, parental education, and home resources) and mathematics achievement.

Keywords Cluster analysis · Instrumentation · Motivation scales · TIMSS sampling

3.1 TIMSS Sampling

TIMSS is an international study of the mathematics and science performance of students at grades four and eight. Starting in 1995, and conducted every four years since then, TIMSS has collected data from multiple countries; more than 60 countries¹ or jurisdictions, and more than 580,000 students participated in the 2015 cycle of assessment. As well as information on mathematics and science performance, the databases include data from context and background questionnaires completed by the students, their teachers, and their parents.

¹The educational systems examined in our analyses are usually referred to as “countries.” This is for ease of reading, but it should be noted that there are a number of systems that are not countries as such, but are units with a degree of educational autonomy that have participated in TIMSS following the same standards for sampling and testing.

In 1995, data were collected from three target populations in 45 countries. These were defined as (a) the two adjacent grades where the majority of nine-year-old students were enrolled, (b) the two adjacent grades where the majority of 13-year-old students were enrolled, and (c) students in their final year of secondary education (Martin and Kelly 1996). In 1999, the target population was limited to grade eight students. From 2003 onwards, the sampling scheme has included students in their fourth and eighth year of schooling, while, in 2015, students in their last year of high school were also sampled.

To select a sample that is representative of the population of students in each participating country, a two-stage random sampling design is used (LaRoche et al. 2016). In the first stage, schools are sampled from each national school sampling frame (a list of all schools with eligible students) with probabilities proportional to school size, and may be stratified by important demographic variables. Once the number of sampled schools is determined in each explicit stratum, systematic sampling proportional to size is used to select schools in each stratum. Provisions for replacement schools are also made. In the second stage, intact classes are chosen through equal probability systematic random sampling. Hence, there is a multilevel structure, where the students are nested in classrooms, and the classrooms nested in schools in each country or jurisdiction. In the 2015 administration, for the sampling precision that is usually required, 150 schools had to be selected for most participating countries, and a sample of 4000 students in each grade (LaRoche et al. 2016).

3.2 Jurisdictions Included in This Study

The countries and benchmarking participants included in the current study were those which participated in the first (1995), the last (2015), and an intermediate administration (2007), provided that they had not been flagged, and that their data are comparable for measuring trends to 2015 (Mullis et al. 2016, Appendix A). The countries which fulfilled these criteria were Australia, England, Hong Kong, Hungary, Iran, Japan, Singapore, Slovenia, and the USA, as well as Norway (grades four and eight), and Ontario and Quebec (Canada), which served as benchmarking participants. These twelve jurisdictions have participated in all TIMSS cycles at both grades four and eight.

Note that, in 1995, data for Ontario and Quebec were obtained as part of a Canadian sample. It was possible to identify schools in those two provinces from the Canada data file using the appropriate school codes (P. Foy, personal communication, 9 August 2018). These two provinces were oversampled in subsequent TIMSS administrations, which makes their results comparable to those of other countries and benchmarking participants.

For the 12 countries and benchmarking participants selected as our sample, the number of grade four students (a fraction of population 1) participating in TIMSS 1995 in each country ranged from 723 to 7296 (see Table 3.1). The range was somewhat smaller for grade eight students (1059 to 7392, a fraction of population 2). Ontario (Canada) had the smallest sample sizes at both grades. All other jurisdictions had much larger sample sizes.

Table 3.1 Sample sizes for the countries and benchmarking participants analyzed (valid percentage of girls)

Participating jurisdictions	TIMSS 1995			TIMSS 2007			TIMSS 2015	
	Population 1 ^a students	Grade 4 students	Population 2 ^a students	Grade 8 students	Grade 4 students	Grade 8 students	Grade 4 students	Grade 8 students
<i>Countries</i>								
Australia	11,248	6065 (49.9)	12,852	7392 (51.4)	4108 (50.0)	4069 (45.3)	6057 (48.9)	10,338 (50.5)
England ^b	6182	3126 (50.6)	3579	1776 (48.0)	4316 (50.0)	4025 (51.8)	4006 (50.6)	4814 (50.7)
Hong Kong	8807	4411 (45.9)	6752	3339 (45.2)	3791 (48.5)	3470 (50.4)	3600 (44.9)	4155 (47.5)
Hungary	6044	3006 (49.8)	5978	2912 (51.1)	4048 (49.7)	4111 (49.9)	5036 (49.8)	4893 (50.6)
Iran	6746	3385 (48.9)	7429	3694 (44.5)	3833 (47.2)	3981 (44.9)	3823 (48.7)	6130 (48.9)
Japan	8612	4306 (50.0)	10,271	5141 (48.5)	4487 (49.3)	4312 (49.7)	4383 (50.2)	4745 (51.0)
Singapore	14,169	7139 (47.4)	8285	4644 (49.7)	5041 (49.2)	4599 (48.8)	6517 (48.8)	6116 (48.7)
Slovenia	5087	2566 (50.5)	5606	2708 (51.1)	4351 (49.5)	4043 (50.0)	4445 (48.4)	4257 (48.2)
USA	11,115	7296 (51.4)	10,973	7087 (50.2)	7896 (51.0)	7377 (50.4)	10,029 (50.6)	10,221 (50.1)
<i>Benchmarking participants</i>								
Norway	4476	– ^c	5736	– ^c	4108 (49.4)	4627 (49.5)	4164 (49.4)	4795 (50.1)
Ontario	1416	723 (45.6)	2078	1059 (49.7)	3496 (49.3)	3448 (50.6)	4574 (48.2)	4520 (49.8)
Quebec	8470	4488 (50.4)	8378	4245 (50.0)	3885 (51.4)	3956 (49.5)	2798 (50.0)	3950 (52.3)

Notes

^aIn 1995 the target populations included: Population 1, defined as “[s]tudents enrolled in the two adjacent grades that contained the largest proportion of 9-year-old students at the time of testing—third- and fourth-grade students in most countries,” and Population 2, which comprised “[s]tudents enrolled in the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing—seventh- and eighth-grade students in most countries” (Gonzales and Smith 1997, pp. 1–4)

^bIn England, year 5 and year 9 are considered equivalent to grades 4 and 8, respectively

^cIn Norway in 1995, only grades 2–3, and grades 6–7 participated, so the two samples were excluded from the analysis

The sample of students participating in TIMSS 2007 was roughly similar in most countries (3448 to 5041 students per jurisdiction) with the exception of the USA, which was the only country with a larger sample of participating students (7896 grade four students and 7377 grade eight students). The number of grade four students from the 12 selected countries participating in TIMSS 2015 ranged from 2798 in Quebec (Canada) to 10,029 in the USA. Finally, the number of grade eight students participating in TIMSS 2015 ranged from 3950 in Quebec (Canada) to 10,221 in the USA. The numbers of students in each sample and administration (see Table 3.1) were sufficient (perhaps with the exception of the Ontario province in the 1995 administration) to allow robust generalizations about populations within each jurisdiction.

3.3 Instrumentation

The TIMSS background questionnaires collect information related to attitudes, motivation, and affect in the study of mathematics. However, there is no solid theoretical background underlying the selection of items that are included in the questionnaires. Reviewing the documentation of the questionnaires, there seems to be a gradual tendency towards a more theoretically-justified selection of items and scales across time. In 1995, the theoretical background did not make reference to motivational theories, and the various items were administered as single indicators; a few items could be grouped into an overall “attitude” scale for both grades (plus a “values” scale for grade eight). In contrast, by the 2015 cycle of TIMSS, the theoretical framework made reference to psychological constructs, including specific motivational variables such as enjoyment, value, and confidence in mathematics; each one was operationalized in a separate, multiple-item scale.

In our analysis, we have taken a construct-level approach: beginning with the latest administration, we extract the relevant motivational variables. Then we attempt to trace items that could represent those constructs in the earlier administrations by relying on TIMSS documentation, item content, and on empirical, factor-analytic evidence. We used exploratory factor analysis, with principal axis estimation and oblique rotation, to validate factor structures where these have not been explicitly reported in the TIMSS manuals. The number of factors was determined by reference to factors having eigenvalues >1 (Kaiser criterion) and by examination of the elbow in scree plots. A factor was accepted if item loadings on the expected factor exceeded 0.30 and cross-loadings were less than 0.30 (Bandalos and Finney 2010).

3.3.1 Motivation Measures in the TIMSS 2015 Administration

Hooper et al. (2013) used Deci and Ryan's (1985) theory of motivation to describe the construct of motivation used in the TIMSS 2015 assessment framework. This theory distinguishes intrinsic from extrinsic motivation, as explained in Chap. 2. In TIMSS 2015, students' enjoyment and interest in learning mathematics was measured with nine four-point Likert-type² items (see Table 3.2). From these nine items a scale variable termed "Students like learning mathematics" was derived, which we also used for our analyses. Confidence in learning mathematics was measured by nine TIMSS questionnaire items, which were used to derive the scale termed "Student confidence in mathematics" (Table 3.2). For grade eight students, nine four-point Likert-type rating items (Table 3.2) were administered to derive a scale to capture the value component. The scale variable derived from these nine items was termed "Students value mathematics" (Hooper et al. 2013).

In the TIMSS 2015 data, context subscales were scaled using the Rasch partial credit item response theory (IRT) model (Masters 1982); corresponding variables are available as described in Martin et al. (2016b). Using the combined data from all participating countries, each item's model parameters were estimated. Subsequently, individual scores for each respondent were computed, ranging from approximately – 5 to 5, and then transformed to a scale that had a mean of 10 and a standard deviation of 2 across all countries. The continuous scales for enjoyment, confidence, and value were used in our analyses.

3.3.2 Motivation Measures in the TIMSS 2007 Administration

In the assessment framework for TIMSS 2007, Mullis et al. (2005) described student motivation as a construct involving students' enjoyment of a subject, values students placed on a particular subject, and their perceived importance of a subject. Student self-concept in mathematics is also considered to influence students' motivation. Student motivation in TIMSS 2007 was measured with seven four-point Likert-type scale items derived from the student background questionnaire (Table 3.3). An additional scale measuring students' value of mathematics, which consisted of four four-point Likert-type scale items, was included in the grade eight student background questionnaire (Foy and Olson 2009).

In TIMSS 2007, items were grouped under three constructs and index variables were calculated for self-confidence (four items), positive affect (three items), and valuing mathematics (four items). However, no scaling was conducted, and unlike the

²In TIMSS, students are asked to indicate their degree of agreement with a number of statements. The response categories for all motivation items are: Agree a lot, Agree a little, Disagree a little, and Disagree a lot.

Table 3.2 TIMSS 2015 questionnaire items used to measure students' enjoyment, confidence, and value

<i>Enjoyment: Students like learning mathematics questionnaire items</i>
(a) I enjoy learning mathematics
(b) I wish I did not have to study mathematics
(c) Mathematics is boring
(d) I learn many interesting things in mathematics
(e) I like mathematics
(f) I like any schoolwork that involves numbers
(g) I like to solve mathematics problems
(h) I look forward to mathematics lessons ("class" instead of "lesson" for grade 8)
(i) Mathematics is one of my favorite subjects
<i>Confidence: Student confidence in mathematics questionnaire items</i>
(a) I usually do well in mathematics
(b) Mathematics is more difficult for me than for many of my classmates
(c) I am just not good at mathematics ("Mathematics is not one of my strengths" for grade 8)
(d) I learn things quickly in mathematics
(e) Mathematics makes me nervous
(f) I am good at working out difficult mathematics problems
(g) My teacher tells me I am good at mathematics
(h) Mathematics is harder for me than any other subject
(i) Mathematics makes me confused
<i>Value: Students value mathematics questionnaire items (grade 8 only)</i>
(a) I think learning mathematics will help me in my daily life
(b) I need mathematics to learn other school subjects
(c) I need to do well in mathematics to get into the <university> of my choice
(d) I need to do well in mathematics to get the job I want
(e) I would like a job that involves using mathematics
(f) It is important to learn about mathematics to get ahead in the world
(g) Learning mathematics will give me more job opportunities when I am an adult
(h) My parents think that it is important that I do well in mathematics
(i) It is important to do well in mathematics

Rasch-scaled variables of the TIMSS 2015 administration, there were no continuous scales for the motivational variables. Therefore, we investigated whether there was empirical support for grouping and averaging items together to create new variables for confidence, enjoyment, and value for mathematics.

At grade four, TIMSS 2007 included motivation scales in the student background questionnaire that were designed to measure their confidence and affect in mathematics (four and three items, respectively; see Table 3.3). We conducted

Table 3.3 TIMSS 2007 questionnaire items used to measure students' confidence, enjoyment, and value

<i>Confidence: Self-confidence in learning mathematics questionnaire items</i>	
(a)	I usually do well in mathematics
(b)	Mathematics is harder for me than for many of my classmates (at grade 8 this was "more difficult" instead of "harder")
(c)	I am just not good at mathematics (at grade 8 this was "mathematics is not one of my strengths")
(d)	I learn things quickly in mathematics
<i>Enjoyment: Students positive affect (enjoyment) toward mathematics questionnaire items</i>	
(a)	I enjoy learning mathematics
(b)	Mathematics is boring
(c)	I like mathematics
<i>Value: Student valuing mathematics questionnaire items (grade 8 only)</i>	
(a)	I think learning mathematics will help me in my daily life
(b)	I need mathematics to learn other school subjects
(c)	I need to do well in mathematics to get into the <university> of my choice
(d)	I need to do well in mathematics to get the job I want

Notes The item "I would like to do more mathematics in school" was administered but not assigned to an index in the TIMSS 2007 framework, so it was not considered in the analysis

exploratory factor analysis (EFA) on each country's sample using principal axis factoring and oblique rotation. In 10 of the samples, two factors were extracted with the Kaiser criterion and explained more than 61% of the variance. The items loaded strongly on their respective factors, with no cross-loadings above 0.30. In two of the samples (Iran and Japan), one factor was extracted using the Kaiser criterion, although the scree plot was ambivalent. Overall, we interpreted this as evidence that these item groups measured two constructs, and that the two sets of items could be combined to create scores for confidence in and enjoyment of mathematics.

We followed a similar approach for grade eight samples. Here, 11 items were included to capture confidence, enjoyment, and value for mathematics (four, three, and four items, respectively). With a principal axis factoring and oblique rotation, nine of the 12 samples resulted in a three-factor solution, as anticipated. At least 63% of the variance was explained by the extracted factors and no cross-loadings above 0.30 were found. In Hong Kong, Iran, and Singapore, two factors were extracted: one comprised the value items and the other comprised of the enjoyment and confidence items. In the Japanese sample, three factors were extracted; however, two of the

value items (“I think learning mathematics will help me in my daily life” and “I need mathematics to learn other school subjects”) cross-loaded on both the value and enjoyment factors.

Since in most of the EFAs item responses loaded onto their intended factors, we created two new variables for grade four and three for grade eight by averaging items (according to the groupings in Table 3.3). If, for an individual student, two or more items responses were missing, we specified the average score as a missing value.

3.3.3 Motivation Measures in the 1995 Administration

No assessment framework existed for TIMSS 1995; however, in the TIMSS 1995 technical report, Martin and Kelly (1996) stated that students’ interest, motivation, and effort were merged in a single construct since they were hard to distinguish from each other. Items that measured students’ values, competence, enjoyment, interest, and importance (five and 12 items in the grade four and eight questionnaires, respectively) were considered to reflect students’ reported motivation for learning mathematics (Table 3.4).

Table 3.4 TIMSS 1995 questionnaire items used to measure students’ confidence, enjoyment, and value

<i>Confidence: Self-confidence in learning mathematics</i>	
(a)	How well do you usually do in math <s> (at grade 8 this was “I usually do well in mathematics”)
(b)	Math <s> is an easy subject (at grade 8 this was “mathematics”)
<i>Enjoyment: Enjoyment of mathematics</i>	
(a)	How much do you like math <s> (at grade 8 this was “mathematics”)
(b)	I enjoy learning math <s> (at grade 8 this was “mathematics”)
(c)	Math <s> is boring (at grade 8 this was “mathematics”)
<i>Value: Value in learning mathematics (included at grade 8 only)</i>	
(a)	I think it is important to do well in mathematics at school
(b)	Mathematics is important to everyone’s life
(c)	I would like a job that involved using mathematics
(d)	I need to do well in mathematics... to get the job I want
(e)	I need to do well in mathematics... to get into the <secondary school> or university I prefer

Notes Items “I need to do well in mathematics... to please my parents” and “I need to do well in mathematics... to please myself” were not included in our calculation of the value in learning mathematics variable

Since the theoretical framework did not describe specific factors beyond general attitudes for mathematics, we conducted parallel analysis for each country dataset to determine the number of motivational factors. Principal axis factoring with oblique rotation and fixed number of factors followed to examine which of the five ordinal items that were administered to grade four students loaded on each factor. For grade four, results showed that items loaded on two motivational factors for all the countries included in the study, except for Iran. Specifically, three items loaded on one factor (i.e., enjoyment) and two on another one (i.e., confidence), based on the examination of the items' contents. Cross-loadings >0.30 were only observed in Hungary, where two of the enjoyment items loaded on both of the factors. In this case, the primary loading coefficients were taken into consideration when determining the factor structure. A single factor was extracted for Iran, and one of the items had a near zero loading.

For grade eight, the findings were more complex. We conducted parallel analysis for each country to determine the number of motivational factors. Results showed that items loaded on four or five motivational factors for most of the countries included in the project (except for Iran). Principal axis factoring with oblique rotation and fixed number of factors followed, to examine which items loaded on each factor. For most countries (except for Iran and Hungary), three items loaded on one factor measuring enjoyment, and two on a confidence factor. In Iran, all five items loaded on a single factor, while in Hungary the two items measuring students' confidence in mathematics loaded onto two different factors. The remaining seven items loaded on two or three factors. The items "I need to do well in mathematics to please my parents" and "I need to do well in mathematics to please myself" formed a factor or were the single items loading on a factor (for seven out of the 11 countries included in our analysis). Because those two items were not included in the TIMSS 2015 administration, we excluded them from further analyses. The remaining five items usually loaded inconsistently onto two factors. Since these items were included in one scale named "value" in TIMSS 2015, we also considered these as forming a single scale in TIMSS 1995 (see Table 3.4). After assessing the results of our EFA, we created variables by averaging items. If for an individual student, two or more item responses were missing, we specified the average score as a missing value.

It is worth noting that, due to local considerations, various nations may not have administered certain items in certain rounds.³ For example, the item "I think it is important to do well in mathematics at school" (Table 3.4) was not administered in Norway in 1995. Nonetheless, the scale score for the construct was calculated for those nations using the remaining items.

³For more detailed information about individual national adaptations undertaken in the TIMSS 2007 and 2015 administrations, see Foy and Olson (2009) and Foy (2017, tables 2 and 3, and supplement 2).

3.4 Other Variables Included in the Study

3.4.1 TIMSS Achievement Score Estimation

To ensure adequate content coverage, a large pool of assessment items is administered in each cycle of TIMSS. The burden of responding to hundreds of questions would be too great for any student, so TIMSS uses a planned missing data, multiple-matrix sampling. Each examinee receives a subset of the item pool. In this way, individual student testing time is reasonable, while it is possible to obtain measures of performance on broad content domains at the aggregate level (Rutkowski et al. 2014). When estimating the distribution of proficiencies for the large population of students, there are problems of inaccuracy in the resulting estimates of individual proficiencies and population characteristics. Plausible values (PVs) are employed to address this problem: all the data, including student responses and their background data are used to estimate PVs. PVs are multiple imputed scores from the estimated conditional ability distributions (given all the students' responses and background data). They can be thought of as imputed scores for "students with similar response patterns and background characteristics in the sampled population" (Martin et al. 2016a, p. 12.5), and aim to present estimates of population parameters; they are not used as estimates of individual student scores. The five PVs estimated for each student are representative of the variation in estimating individual student proficiency (Martin et al. 2016a).

Achievement data were scaled through IRT models, which are latent variable models that estimate the probability of a specific student answering items correctly based on the student's proficiency, which is the latent trait θ . For dichotomous response questions (e.g., multiple-choice items marked as correct or incorrect) the three-parameter IRT model was used; this model accounts for three parameters: item difficulty, item discrimination, and item pseudo-chance. For constructed response items, which do not have options from which to select but which are also marked as correct or incorrect, a two-parameter IRT model with parameters for item discrimination and difficulty was employed. For polytomous items, the partial credit model was used (Martin et al. 2016a).

International achievement in TIMSS 1995 was reported using only one plausible value, however all five PVs are available in the TIMSS 1995 data (Gonzales and Smith 1997). TIMSS cycles after 1995 reported five PVs and we used these as indicators of students' achievement in mathematics. When we report the average performance for a student group, all five PVs were considered and total student weights were applied using the IEA's International Database Analyzer (IDB) software (see www.iea.nl/data for further information about this free-to-download analysis tool).

3.4.2 *Other Variables of Interest*

Student Sex The students' sex variable was used in the present study.

Time on Homework Items measuring the number of hours spent on studying or doing homework differ across TIMSS administrations. TIMSS 1995 student questionnaire included the question "On a normal school day, how much time before or after school do you spend doing each of these things? [...] studying math or doing math homework after school," which measured the time a student spent on studying mathematics. Two questions examining the time spent on mathematics homework ("How often does your teacher give you homework in mathematics?" and "When your teacher gives you mathematics homework, about how many minutes do you usually spend on your homework?") were included in the TIMSS 2007 and TIMSS 2015 student questionnaires. An index variable, consisting of three categories, combined the two responses in the 2007 dataset. A derived variable with three categories was extracted from the two items; this measured the weekly time a student spent on mathematics homework. The time spent on mathematics homework item was not included in the TIMSS 2015 grade four student questionnaire. All items measuring time spent on studying/doing homework in the three TIMSS administrations that we considered (1995, 2007, and 2015) contained five response options.

Parental Education Across all TIMSS administrations, the grade eight student questionnaire includes questions asking students to report the highest level of parental education. Parental education questions were omitted from the grade four student questionnaires. The items measuring parental educational level are highly similar across TIMSS administrations, however the response options vary across administrations and across nations. At grade eight, in all three of the TIMSS administrations that we selected for our study, two separate items asked for the highest education level achieved by mother and father. There were eight response options in 2015 and 2007, and just seven response options in 1995. A derived variable with six categories of education levels in 2015 and 2007, and four categories of education level in 1995, was created by combining the observed variables about the father and mother into a parental education variable. Here we report the percentage of parents with education above a cut point: "above secondary" for TIMSS 1995 and "post-secondary and above" for TIMSS 2007. We did not analyze this variable for TIMSS 2015 because it was included in the definition of the more comprehensive home resources variables.

Home resources Home resources are proxy variables for a student's socioeconomic status (SES), and are only available in the TIMSS 2015 data. Items for parental education level, occupation, income, and number of books in home are used as indicators of students' SES. There are two derived scale variables in TIMSS 2015: the "Home resources for learning" scale for grade four students and the "Home educational resources" scale for grade eight students. The two derived scales include number of books and number of children's books at home, number of home study supports (own room and internet connection) and parental educational and

occupational level (grade four) or number of books in the home, number of home study supports (own room and internet connection), and parental educational level (grade eight) (Martin et al. 2016b).

These two scales were calculated for TIMSS 2015, but not for TIMSS 2007 and 1995, because relevant items have not been surveyed consistently. Hence, we were unable to include proxies for SES in the analyses of earlier administrations.

3.5 Analysis Technique

In cluster analysis, similar observations in a dataset are grouped together in a cluster (Bartholomew et al. 2008). Similarity is determined by information from one or more of the variable characteristics of the observations. The grouping is not known in advance. Identification of homogenous observations is essentially a taxonomy analysis.

While there are several techniques for cluster analysis, we here outline three common approaches that are available in statistical packages (such as the IBM SPSS Statistics package; see https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/base/cluster_choosing.html). Firstly, hierarchical cluster analysis is an agglomerative procedure that begins with each observation as a separate group, and gradually combines observations or groups based on similarity, until one large cluster is formed. The hierarchical approach is recommended when input variables are continuous and the sample of observations is small. A dendrogram is produced and examined to ascertain the number of clusters to retain and their meaning. K-means clustering can be used with continuous variables and large datasets. The number of clusters must be defined in advanced. Solutions with different numbers of clusters can be inspected and compared. Finally, two-step cluster analysis can handle both continuous and categorical variables in very large datasets to generate a solution; first by constructing a cluster features tree to summarize the observations and then by employing an agglomerative algorithm.

Because cluster analysis is an exploratory procedure, different numbers of clusters may be extracted and interpreted, especially when using two-step or K-means clustering. In our preliminary analyses, a small number of clusters were extracted (e.g., two or three). In these solutions, the clusters were consistent and not very informative with respect to the input variables. For example, one cluster was composed of students with high scores on all input variables, another cluster grouped the students with moderate scores, and a third cluster was composed of students with rather low motivation scores. This approach did not permit the identification of possible inconsistent profiles across the motivational constructs, which was an important aim of our study.

Therefore, within a two-step cluster approach, the fixed number of clusters was incremented between three and five at grade four and between three and six at grade eight; more clusters were examined for grade eight because one additional input variable (“Value for mathematics”) was available for these older students. These

numbers were selected so that (a) the analysis would produce more than just clusters with consistent motivation responses, (b) a manageable number of reasonably-sized clusters would be produced, increasing the likelihood they would cross-validate, and (c) inconsistent motivational profiles could be identified. Clusters with students scoring medium-to-low on one input variable and high on another input variable present potential theoretically interesting opportunities. For instance, students who value, but do not necessarily like mathematics, or students with high self-confidence in the subject, but who report low value and enjoyment in mathematics classes, may have more or less successful achievement profiles or differ on sociodemographic characteristics. Cluster comparisons on evaluation variables such as achievement and demographics offer insights into the possible predictors or outcomes of such inconsistent motivational profiles.

Due to the exploratory nature of cluster analysis, the evaluation of competing cluster solutions was not automatically determined. In choosing the final solution, we considered statistical measures, such as the silhouette measure of cohesion and separation (at least “fair”; Kaufman and Rousseeuw 1990), and the relative size of the smallest cluster (>7% of the sample). In addition, we considered the interpretation of the derived clusters. The final number of clusters for each country sample, in each cycle of TIMSS (2015, 2007, and 1995), and at each grade (four and eight) was decided based on the assessment of two independent researchers. When agreement could not be reached, a decision was adjudicated in the presence of a third researcher.

The two-step cluster method was implemented for each sample. This approach is available in the IBM SPSS Statistics software and is appropriate for large samples (Appendix A provides the syntax used to generate clusters). The following measures were entered as input variables:

- (1) Students like learning mathematics/enjoyment
- (2) Students confident in mathematics/confidence
- (3) Students value mathematics/value (this applies only for grade eight).

These scale variables were available in the TIMSS 2015 administration datasets and were derived from context questionnaire items using IRT procedures. For the TIMSS 2007 and 1995 administrations, the variables used were derived after theoretical considerations and factor analytic procedures, by averaging the relevant items as described in the instrumentation section. Hence, the motivation scores in TIMSS 2015 are expressed on a different scale from that used for the earlier administrations.

Because the results of the clustering algorithm in SPSS depend on the order of cases in a dataset, prior to each analysis we undertook the following steps: (a) the cases of each dataset were sorted by each student’s unique ID, (b) a fixed seed was specified, and (c) an observation was generated from a continuous uniform distribution for each case for a random ordering of the dataset. Thus, each cluster solution was based on the same reproducible ordering of cases. The following code was always added before any clustering procedure:

```
sort cases by IDSTUD(A).  
set rng mc seed 123456789.  
compute randvar=rv.uniform(1,1000).
```

sort cases by randvar.
delete variables randvar.

Because the TIMSS data has a nested structure, we here note that the literature identifies two approaches for dealing with sampling weights. The design-based approach recommends using the sampling weights in order to avoid biased parameter estimates. Conversely, the model-based approach does not suggest the use of sampling weights because, if the correct (“true”) model is specified, the use of weights leads to a decrease in efficiency and precision (Anderson et al. 2014; Snijders and Bosker 2012). The IBM SPSS Statistics two-step cluster procedure also does not permit the use of sampling weights and ignores the specification on the WEIGHT command.⁴ Thus, we did not use sampling weights in our cluster analyses. With respect to missing values, cases were excluded from the cluster analysis when a value was missing from the input variables.

When evaluating the clusters, we examined the following background and achievement variables for each cluster:

- (1) Average performance in mathematics (PVs 1 through 5)
- (2) Percentage of girls in the cluster
- (3) Percentage of students with a high level of parental education (applies only for grade eight students, and administrations 2007 and 1995)
- (4) Home resources: the average “home resources for learning” (only available for TIMSS 2015 grade four students) and “home educational resources” (only available for TIMSS 2015 grade eight students), as indications of SES
- (5) Time spent on homework, with the caveats:
 - TIMSS 2015: This was not available for grade four students; at grade eight, TIMSS 2015 reported the percentage of students that spent “> 45 min on homework weekly.”
 - TIMSS 2007: The variable used was the “index of time on math homework.”
 - TIMSS 1995: TIMSS 1995 reported the percentage of students that spent “more than 1 h on homework daily.”

Finally, we conducted statistical tests for comparing cluster means. We undertook pairwise mean comparisons using the IEA’s IDB Analyzer, which allowed us to estimate weighted statistics and corrected standard errors for all the TIMSS assessments. Clusters were compared on average performance in mathematics using all five plausible values for all administrations and samples, and, for 2015, on the home resources variables. Since multiple pairwise tests were conducted for each jurisdiction, we adopted an alpha level of 0.001; a difference was considered statistically significant if the t-statistic (in absolute value) exceeded 3.29. We employed chi-square testing to examine dependencies between gender and clusters. Parental education and homework engagement were measured with different response scales across samples; we report descriptive statistics by cluster for those two variables.

⁴https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/base/syn_twostep_cluster_overview.html.

References

- Anderson, C. J., Kim, J. S., & Keller, B. (2014). Multilevel modeling of categorical response variables. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 481–519). Boca Raton, FL: CRC Press.
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). New York, NY: Routledge.
- Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. Boca Raton, FL: Chapman and Hall/CRC.
- Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Foy, P. (2017). *TIMSS 2015 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UserGuide.pdf.
- Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Gonzales, E. J., & Smith, T. A. (1997). *User guide for the TIMSS international database. Primary and Middle School Years*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 context questionnaire framework. In I. V. S. Mullis, & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 61–82). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/timss2015/frameworks.html>.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York, NY: John Wiley and Sons.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sample design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 3.1–3.37). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Martin, M. O., & Kelly, D. L. (Eds.). (1996). *Third International Mathematics and Science Study (TIMSS) technical report, Volume I: Design and development*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College. Retrieved from <https://timss.bc.edu/timss1995i/TechVol1.html>.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (2016a). TIMSS 2015 achievement scaling methodology. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 12.1–12.9). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., & Palazzo, L. (2016b). Creating and interpreting the TIMSS 2015 context questionnaire scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/timss2015/international-results/>.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <https://timss.bc.edu/TIMSS2007/frameworks.html>.

- Rutkowski, L., Gonzales, E., von Davier, M., & Zhou, Y. (2014). Assessment design for international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 75–95). Boca Raton, FL: CRC Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis* (Second ed.). London, UK: Sage.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

Cluster Analysis Results for TIMSS 2015 Mathematics Motivation by Grade and Jurisdiction



Abstract A person-centered approach can be used to identify the motivational profiles of grade four and grade eight students participating in successive cycles of IEA's Trends in International Mathematics and Science Study (TIMSS); such analyses may be a powerful route to developing a better understanding of student motivation patterns and their interaction with achievement and other background variables. This chapter presents results for TIMSS 2015, which provided the most comprehensive motivational data for measuring students' enjoyment of, confidence in, and value for mathematics. A two-step cluster approach was applied in each of the 12 jurisdictions, and at both grades, illustrating in detail the techniques applied to all three TIMSS administrations. Visual inspection of variable distributions by cluster, and descriptive and inferential statistics across diverse samples highlight some cross-culturally robust patterns. Consistent with variable-centered findings, clusters that had consistently high scores for all motivational variables outperformed those with consistently low motivation scores on the TIMSS mathematics achievement test. However, clusters with inconsistent motivational profiles tended to have higher mean mathematics score when students' confidence in their ability to do mathematics was strong.

Keywords Cluster differences • Consistent profiles • Inconsistent profiles • Motivation clusters • Person-centered approach

4.1 Introducing the Person-Centered Approach

We used a person-centered approach to study the motivational profiles of the grade four and grade eight students participating in TIMSS. Our two-step cluster analysis used the variables enjoyment of, confidence in, and value for mathematics (this last variable was available only for grade eight). As an exploratory procedure, various solutions with between three and six clusters (between three and five for grade four) were examined for each sample. Chapter 3 reviews the judgement criteria for deciding the number of extracted clusters. We repeated the analysis for all 12 jurisdictions, for both grades, for the TIMSS 1995, 2007, and 2015 administrations.

In this chapter, we present results by jurisdiction for the TIMSS 2015 grade four and grade eight samples in alphabetical order by country. In TIMSS 2015, motivation scales contained more items than earlier administrations (see Chap. 3). Scale scores were derived after advanced latent variable methodology was completed for the 2015 motivation variables and for the home resources measure (latent variable scores had been unavailable in previous administrations). In the TIMSS 2007 and 1995 cycles, simpler procedures were used for scoring: averages of items were estimated to generate the motivation variables, and single items were used as proxies for socioeconomic background. The distributions of motivation variables by cluster and jurisdiction, and descriptive statistics with statistical tests by cluster and jurisdiction for the TIMSS 1995 and 2007 administrations are provided in Appendices B and C. We present summaries of the cluster analysis results for all administrations and grades, along with trend comparisons, in Chap. 5.

4.2 Cluster Analysis Results for the TIMSS 2015 Administration at Grade Four by Jurisdiction

Grade four students responded to two motivation scales in the TIMSS 2015 administration: (1) confidence, and (2) enjoyment in mathematics. Scores on the two scales were cluster analyzed, and solutions of three to five clusters were examined in each sample. Based on the criteria presented in Chap. 3, a solution was chosen.

We here present a detailed description of the clusters that were produced in relation to the grade four TIMSS 2015 mathematics results for each country. Boxplots depict the distribution of scores for the two motivational variables within each cluster, and the width of each boxplot represents the size of the cluster. Descriptive statistics by cluster are presented for cluster size, mean achievement, gender composition, and mean home resources for learning. In all tables presented in this chapter (and Appendix C), the procedure “Percentages and Means” in the IEA IDB Analyzer was used to obtain weighted means by cluster for PVs and for home resources (using student weight).

4.2.1 *Australia*

Four clusters were identified in the Australian sample (Fig. 4.1). A small cluster of students (cluster 1) responded very strongly on confidence and strongly on enjoyment; the second cluster had high enjoyment ratings, while the remaining clusters had consistently moderate or low distributions of enjoyment and confidence. Mean achievement differed significantly across most groups, with the highest performance appearing in the cluster with the highest confidence scores (Table 4.1). Performance was lower in clusters with lower motivation scores. There were fewer girls in the more motivated and better performing clusters. Although relatively high

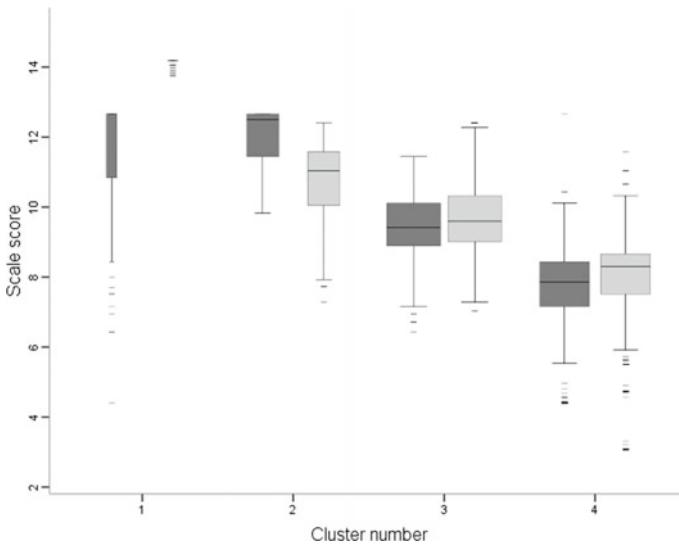


Fig. 4.1 Distributions of motivation variables by cluster for Australia, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.1 Descriptive statistics by cluster for Australia, TIMSS 2015 grade 4

Cluster characteristics	Cluster number			
	1	2	3	4
Size (%)	7.0	21.9	36.9	34.1
Mean plausible value	581.0 ^a	536.1 ^b	524.3 ^b	485.8 ^c
Female students in cluster (%) [*]	32.1	42.3	48.8	56.8
Mean home resources for learning scale score	12.0 ^a	11.4 ^{a,b}	11.5 ^{a,b}	11.3 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 121.109, p < 0.001$)

for all Australian clusters, fewer home resources for learning were found in the lower motivation groups; differences were not always statistically significant among groups.

4.2.2 Canada-Ontario

Cluster analysis resulted in five clusters for Ontario (Fig. 4.2). The smallest cluster (cluster 5) had very high confidence and high enjoyment distributions, and had the highest mean achievement. Comparison of clusters 3 and 4, which both had moderate

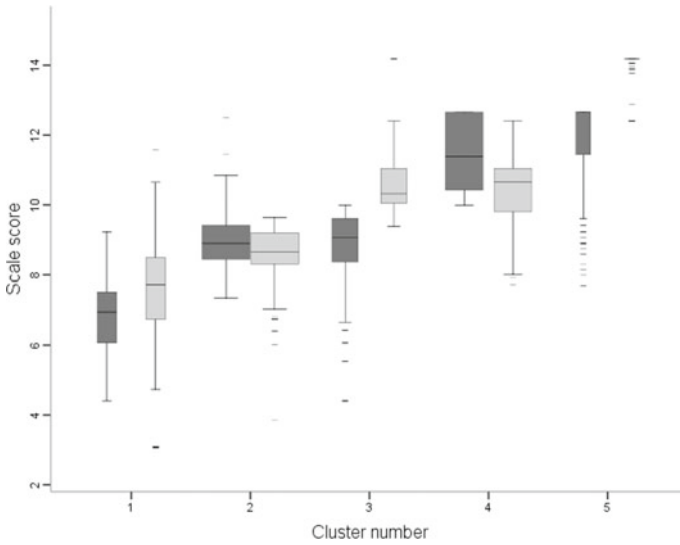


Fig. 4.2 Distributions of motivation variables by cluster for Ontario, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

confidence endorsements, revealed that achievement was higher for cluster 3, even though enjoyment was markedly lower compared to cluster 4. Students in cluster 2 reported similar levels of enjoyment to those in cluster 3, but had lower confidence scores, and significantly lower achievement scores. There were more boys in the more motivated clusters, while the level of home resources for learning was significantly lower in the lower performing clusters (Table 4.2).

Table 4.2 Descriptive statistics by cluster for Ontario, TIMSS 2015 grade 4

Cluster characteristics	Cluster number				
	1	2	3	4	5
Size (%)	13.3	32.9	17.7	25.9	10.2
Mean plausible value	479.9 ^a	485.5 ^a	542.5 ^b	525.8 ^c	564.8 ^d
Female students in cluster (%) [*]	52.9	53.7	46.0	43.8	40.2
Mean home resources for learning scale score	11.1 ^{a,b}	11.0 ^a	11.5 ^{c,d}	11.3 ^{b,c}	11.6 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 45.268, p < 0.001$)

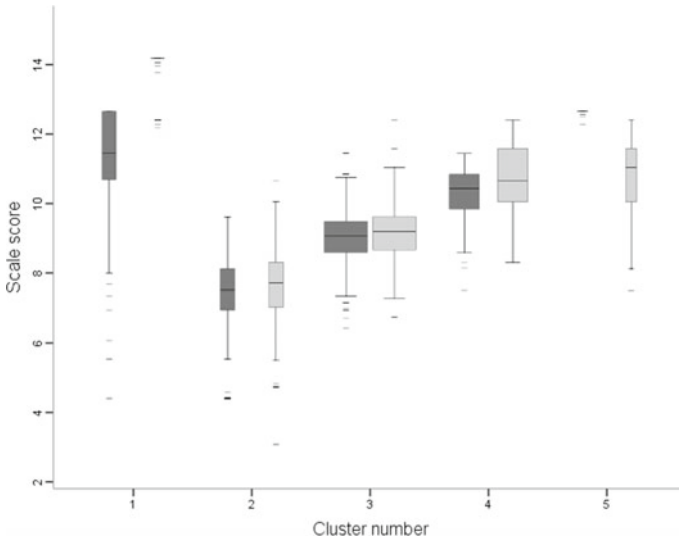


Fig. 4.3 Distributions of motivation variables by cluster for Quebec, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.3 Descriptive statistics by cluster for Quebec, TIMSS 2015 grade 4

Cluster characteristics	Cluster number				
	1	2	3	4	5
Size (% of total number of students)	12.3	12.7	38.5	26.7	9.8
Mean plausible value	577.5 ^a	498.0 ^c	520.8 ^d	552.4 ^b	544.2 ^b
Female students in cluster (%)*	36.6	63.6	55.9	45.4	38.4
Mean home resources for learning scale score	11.3 ^{a,b}	10.8 ^{a,b}	10.9 ^b	11.2 ^a	11.0 ^{a,b}

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 84.995, p < 0.001$)

4.2.3 Canada-Quebec

Five clusters were extracted from the Quebec sample (Fig. 4.3). Cluster 1 had very high confidence and high enjoyment score distributions, and had the highest mean achievement score. Clusters 4 and 5 had similar confidence score distributions and mean achievement did not differ significantly, despite the difference in enjoyment of mathematics between these two clusters. Clusters 2 and 3 had lower motivation distributions and lower achievement scores. Boys were overrepresented in the three higher motivation clusters, and girls overrepresented in the lower motivation clusters

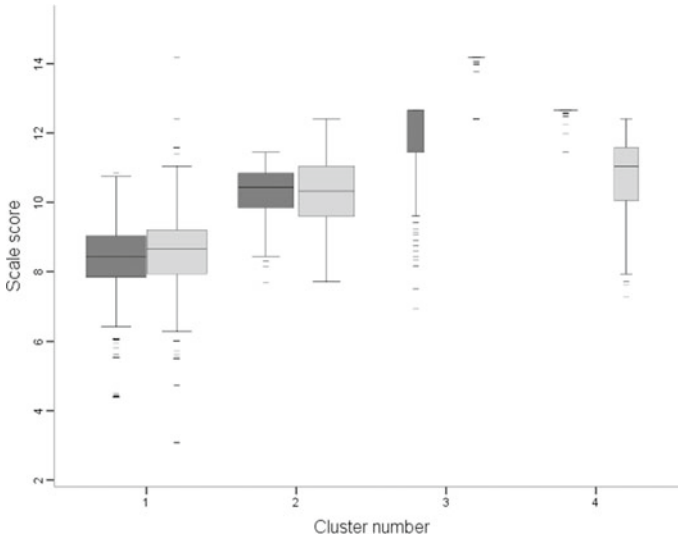


Fig. 4.4 Distributions of motivation variables by cluster for England, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

(Table 4.3). Differences in levels of home resources for learning were small and not always statistically significant.

4.2.4 *England*

Of the four clusters found in the English sample, the smallest in size (cluster 3) had very high endorsements of confidence and high scores on enjoyment; it was also the cluster with the highest mean achievement (Fig. 4.4). Comparing clusters 2 and 4, which had similar distributions on confidence, cluster 2 had lower endorsements of enjoyment, and higher (but not significantly) mean achievement score. There were more boys in the high motivation groups and more girls in the lower performing groups (Table 4.4). The variable home resources for learning was not available for the English sample.

4.2.5 *Hong Kong*

Five clusters were identified in Hong Kong (Fig. 4.5). Confidence and enjoyment distributions were fairly consistent within most clusters, except in cluster 1, which was the best for mean achievement, and cluster 3, which had high endorsement for enjoyment. Mean achievement was positively associated with confidence scores at the

Table 4.4 Descriptive statistics by cluster for England, TIMSS 2015 grade 4

Cluster characteristics	Cluster number			
	1	2	3	4
Size (% of total number of students)	38.4	35.5	10.7	15.5
Mean plausible value	524.1 ^a	560.5 ^b	582.1 ^c	544.3 ^b
Female students in cluster (%)*	58.3	51.5	34.0	42.2

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 97.897, p < 0.001$)

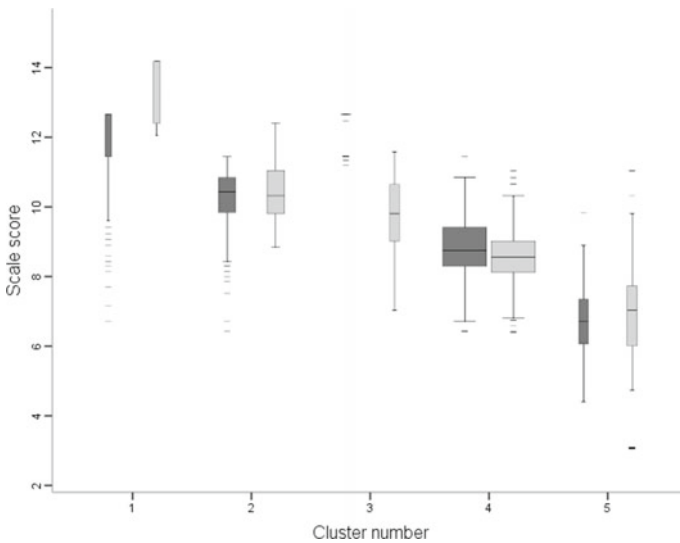


Fig. 4.5 Distributions of motivation variables by cluster for Hong Kong, TIMSS 2015 grade 4

Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

cluster level. Cluster 4 comprised half of the sample, and it was balanced with respect to gender (Table 4.5). The three higher motivation clusters were predominantly male, while cluster 5 contained more girls than boys. The levels of home resources for learning were slightly higher in the higher motivation groups, but differences were small.

Table 4.5 Descriptive statistics by cluster for Hong Kong, TIMSS 2015 grade 4

Cluster characteristics	Cluster number				
	1	2	3	4	5
Size (% of total number of students)	7.6	20.0	11.2	50.4	11.0
Mean plausible value	665.8 ^a	646.3 ^a	619.1 ^b	600.9 ^c	581.6 ^d
Female students in cluster (%)*	25.9	37.0	31.3	51.3	58.2
Mean home resources for learning scale score	10.9 ^a	10.5 ^{a,b}	10.2 ^{b,c}	10.1 ^c	10.3 ^{a,b,c}

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 144.738, p < 0.001$)

4.2.6 Hungary

Five clusters were extracted from the Hungarian sample (Fig. 4.6). When the motivation variable distributions were high, mean achievement was also high. Clusters 2 and 4 had similar confidence scores, but mean achievement was higher in cluster 2, despite enjoyment scores being lower than in cluster 4. There were more boys than girls in the high motivation and high achievement clusters (Table 4.6). Significant differences were identified for the home resources for learning variable, with higher levels for this variable being associated with the higher achieving clusters.

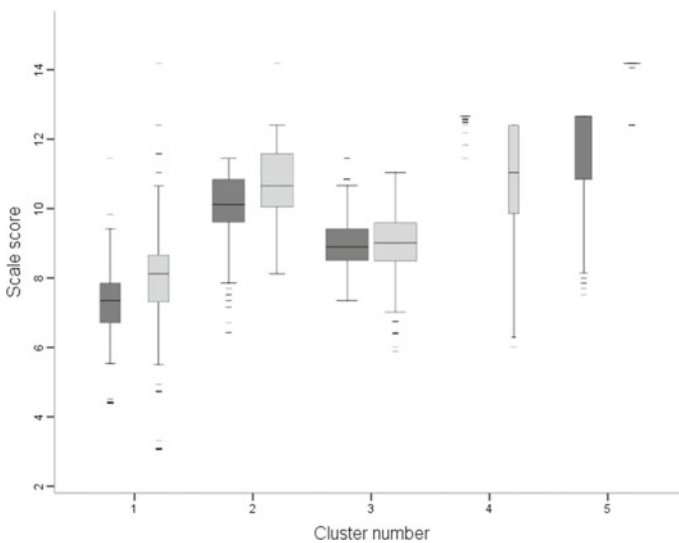


Fig. 4.6 Distributions of motivation variables by cluster for Hungary, TIMSS 2015 grade 4

Notes Dark grey = enjoyment, light grey = confidence. Box width represents relative cluster size

Table 4.6 Descriptive statistics by cluster for Hungary, TIMSS 2015 grade 4

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	16.6	26.3	34.7	8.6	13.8
Mean plausible value	486.7 ^a	557.0 ^b	505.5 ^c	533.4 ^d	590.4 ^e
Female students in cluster (%)*	55.7	47.8	54.5	41.8	39.2
Mean home resources for learning scale score	10.3 ^{a,b}	10.6 ^b	10.1 ^a	10.0 ^a	11.0 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 71.670, p < 0.001$)

4.2.7 Iran

Of the five clusters extracted from the Iranian sample, cluster 5 had high scores on the two motivation variables and high mean achievement (Fig. 4.7). Clusters 3 and 4 had similar distributions for confidence and similar mean achievement, despite different levels of enjoyment for mathematics. The same was true for clusters 1 and 2, which contained students who had lower levels of motivation and achievement. Gender composition was not significantly different across clusters in the Iranian

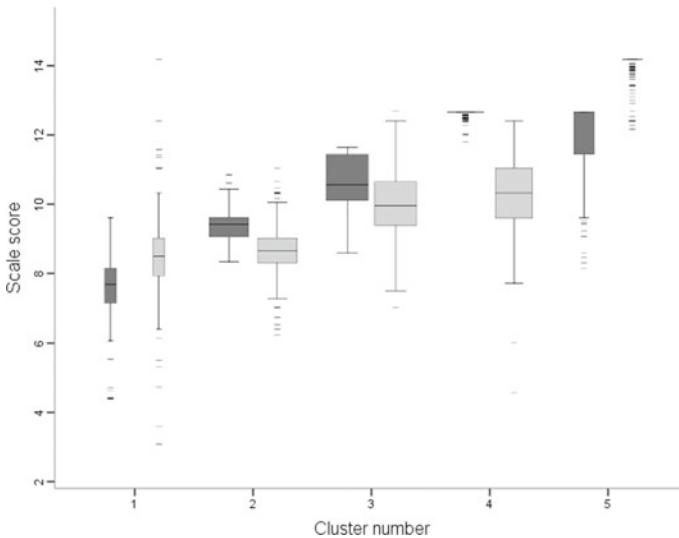


Fig. 4.7 Distributions of motivation variables by cluster for Iran, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.7 Descriptive statistics by cluster for Iran, TIMSS 2015 grade 4

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	7.5	26.4	28.1	24.5	13.5
Mean plausible value	398.6 ^{a,b}	386.1 ^a	435.4 ^{b,c}	440.8 ^{c,d}	463.2 ^d
Female students in cluster (%)*	44.3	49.5	48.1	49.2	52.4
Mean home resources for learning scale score	8.7 ^a	8.1 ^b	8.3 ^{a,b}	8.3 ^{a,b}	8.5 ^{a,b}

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was not significant ($\chi^2(4) = 5.239, p = 0.264$)

sample (Table 4.7). Differences in the home resources for learning variables were small across the clusters.

4.2.8 Japan

A consistent pattern of motivation score distributions was found across the four Japanese clusters (Fig. 4.8). Clusters with students who reported higher confidence and enjoyment of mathematics had significantly higher mean achievement, more

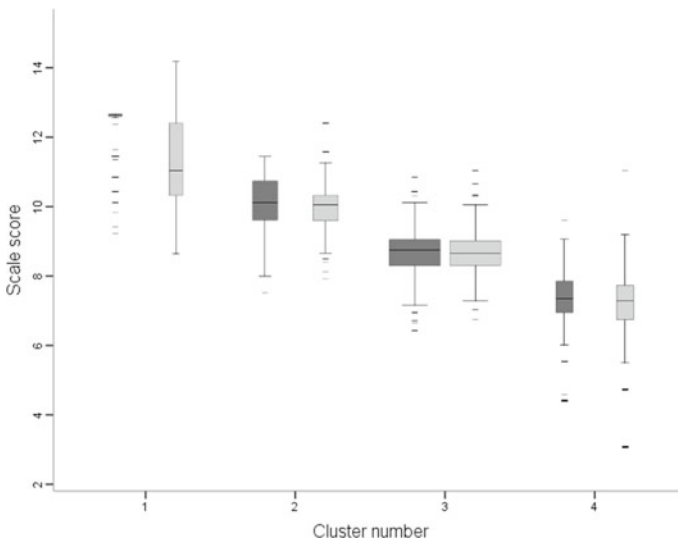


Fig. 4.8 Distributions of motivation variables by cluster for Japan, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.8 Descriptive statistics by cluster for Japan, TIMSS 2015 grade 4

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	12.7	23.5	47.8	15.9
Mean plausible value	636.8 ^a	616.8 ^b	583.6 ^c	552.2 ^d
Female students in cluster (%) [*]	40.4	44.4	54.9	52.3
Mean home resources for learning scale score	10.7 ^a	10.4 ^b	10.1 ^c	9.9 ^d

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(3) = 55.309, p < 0.001$)

home resources for learning, and contained more boys than girls compared to the clusters reporting lower motivation (Table 4.8).

4.2.9 Norway

Five clusters were identified in the Norwegian sample (Fig. 4.9). Cluster 1 had very high confidence and high enjoyment scores, and had the highest mean achievement; cluster 5 had the lowest motivation score distributions and the lowest mean

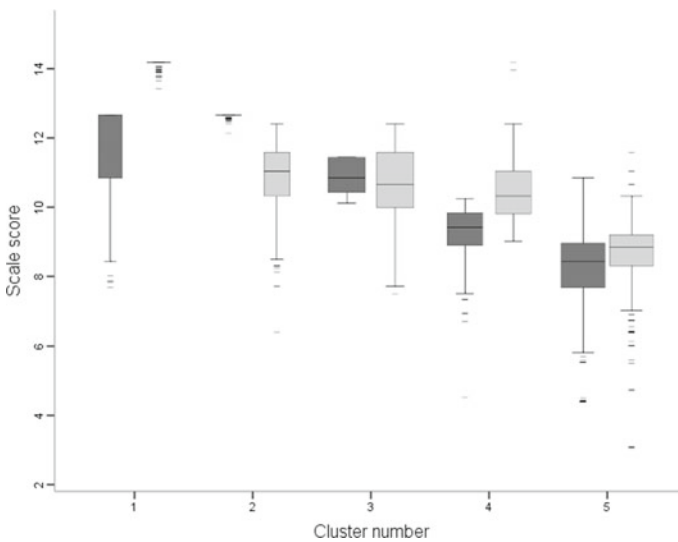


Fig. 4.9 Distributions of motivation variables by cluster for Norway, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.9 Descriptive statistics by cluster for Norway, TIMSS 2015 grade 4

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	14.3	15.5	22.4	21.3	26.4
Mean plausible value	527.3 ^a	492.3 ^b	499.4 ^b	505.2 ^b	462.5 ^c
Female students in cluster (%)*	42.2	51.8	51.2	47.5	52.3
Mean home resources for learning scale score	11.5 ^{a,b}	11.4 ^{a,b}	11.5 ^{a,b}	11.7 ^a	11.3 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 19.689, p = 0.001$)

achievement. For the remaining three clusters, the confidence variable distributions were largely overlapping, and scores for enjoyment were the distinguishing characteristic: enjoyment was high in cluster 2, moderate in cluster 3, and moderate-to-low in cluster 4. Nevertheless, none of the clusters differed significantly in terms of achievement (Table 4.9). A larger percentage of boys than girls was found in the highest motivation cluster, but gender differences were otherwise small. Differences in the home resources for learning variable were also small across clusters and not always statistically significant.

4.2.10 Singapore

Four clusters were identified in the Singapore sample (Fig. 4.10). Students with above average endorsement of confidence and enjoyment (cluster 4) had the highest mean achievement. Clusters 1 and 3 contained students reporting moderate or lower confidence in mathematics, and had similar mean achievement, despite higher endorsement of enjoyment in cluster 1. Cluster 2 contained students with the lowest scores for the motivation variables, the lowest mean achievement, and more girls than boys. In contrast, there were more boys than girls represented in the highest confidence cluster (cluster 4) (Table 4.10). Across the clusters, there were significant differences in the home resources for learning variable, with higher scores reported for the higher achieving clusters.

4.2.11 Slovenia

Five clusters were identified in the Slovenian sample (Fig. 4.11). Cluster 1 contained students with very high confidence and moderate-to-high enjoyment scores, and had the highest mean achievement among the clusters. Clusters 2 and 3 contained students

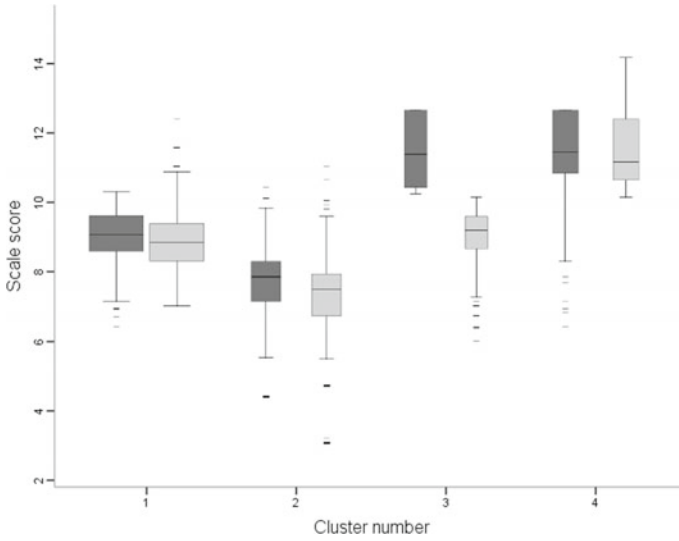


Fig. 4.10 Distributions of motivation variables by cluster for Singapore, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.10 Descriptive statistics by cluster for Singapore, TIMSS 2015 grade 4

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	41.2	22.2	17.1	19.5
Mean plausible value	616.6 ^a	575.1 ^b	605.7 ^a	676.6 ^c
Female students in cluster (%) [*]	50.8	54.7	50.1	36.9
Mean home resources for learning scale score	10.8 ^a	10.5 ^b	10.6 ^b	11.4 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 96.964, p < 0.001$)

with similar mean achievement, although motivation scores (particularly enjoyment in mathematics) were lower in cluster 3. The other two clusters had lower distributions of the motivation variables and lower mean achievement. There were more boys than girls in the higher motivation clusters (Table 4.11). Across the clusters, there were small differences in mean home resources for learning; clusters with higher motivation had higher levels for this variable.

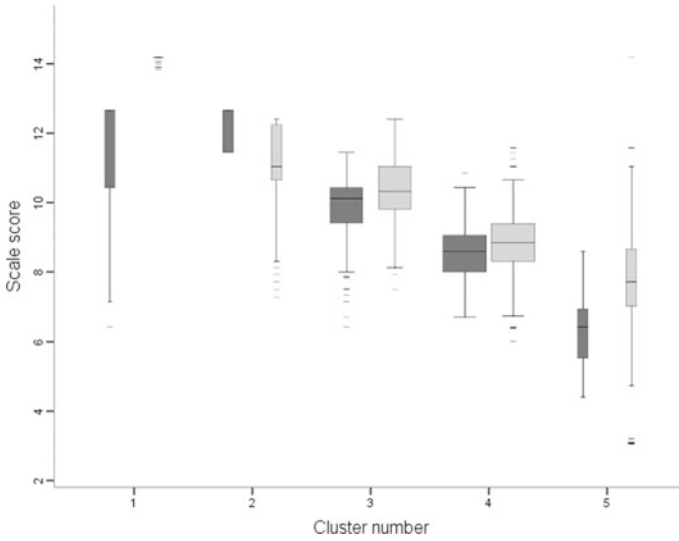


Fig. 4.11 Distributions of motivation variables by cluster for Slovenia, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.11 Descriptive statistics by cluster for Slovenia, TIMSS 2015 grade 4

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	9.2	10.1	30.4	40.8	9.5
Mean plausible value	569.0 ^a	532.5 ^b	536.9 ^b	502.2 ^c	487.1 ^d
Female students in cluster (%)*	34.0	39.5	47.9	54.5	49.9
Mean home resources for learning scale score	11.1 ^a	10.6 ^{b,c}	10.8 ^{a,b,d}	10.5 ^c	10.6 ^{c,d}

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|\mathit{t}| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 74.794, p < 0.001$)

4.2.12 USA

Of the five clusters identified in the USA sample, cluster 1, with very high confidence and moderate-to-high enjoyment scores, had a high mean achievement (Fig. 4.12). Cluster 2 also contained students with high levels of mean performance, ahead of cluster 3, which included students with strong endorsement of enjoyment in mathematics. The remaining two clusters had consistent below moderate (cluster 4) and low (cluster 5) motivation scores, and did not significantly differ in achievement. There was a significant difference in gender composition across the clusters, with

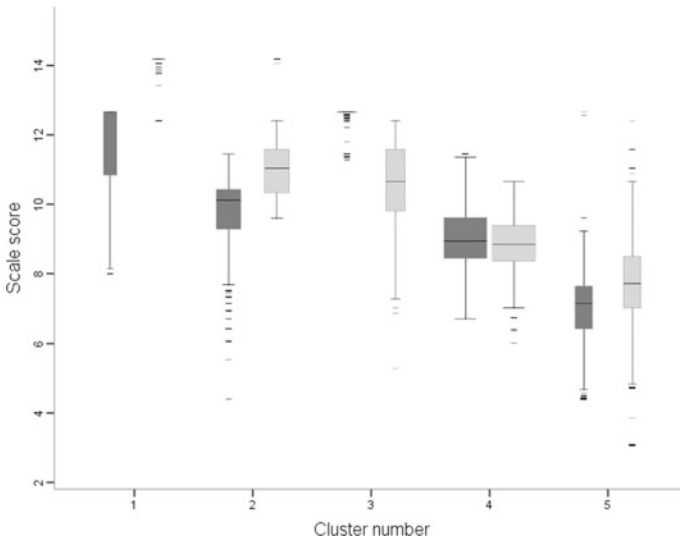


Fig. 4.12 Distributions of motivation variables by cluster for USA, TIMSS 2015 grade 4
Notes Dark gray = enjoyment, light gray = confidence. Box width represents relative cluster size

Table 4.12 Descriptive statistics by cluster for USA, TIMSS 2015 grade 4

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	11.3	20.9	16.4	36.6	14.8
Mean plausible value	589.4 ^a	580.6 ^a	544.8 ^b	515.5 ^c	509.2 ^c
Female students in cluster (%) [*]	43.2	46.2	51.4	54.4	55.6

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $tl > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 73.133, p < 0.001$)

more boys than girls in the two highest confidence clusters, and more girls than boys in the two lower motivation clusters (Table 4.12). The home resources for learning variable was not available for the USA sample.

4.3 Cluster Analysis Results for the TIMSS 2015 Administration at Grade Eight by Jurisdiction

Grade eight students responded to three motivation scales in the TIMSS 2015 administration: enjoyment of, confidence in, and value for mathematics. We performed cluster analyses of the scores on the three scales, and explored solutions

providing between three and six clusters for each of the 12 samples. Following the criteria we outlined in Chap. 3, a solution was selected.

We here present a detailed description of the clusters that were generated from the grade eight TIMSS 2015 mathematics results for each country. Boxplots depict the distribution of scores for the three motivational variables for each cluster, and the width of each boxplot represents the size of the cluster. Descriptive statistics by cluster are presented for cluster size, mean achievement, gender composition, homework engagement, and mean home educational resources.

4.3.1 Australia

Five clusters were extracted from the Australian sample (Fig. 4.13). The two smaller clusters consist of students who reported high scores on all variables (cluster 1) or low scores for all variables (cluster 5). The other three clusters were larger in size. Clusters 2 and 3 had similar distributions for confidence, but cluster 3 reported slightly higher scores for enjoyment and much higher value scores. Cluster 4 had moderate to low scores for all three variables.

Comparison of mean achievement for the clusters (Table 4.13) revealed significant differences: when students' motivational scores were higher, their mean achievement was also significantly higher. An interesting finding is that clusters 2 and 3 did not differ in their mean achievement scores, despite the large difference in value and

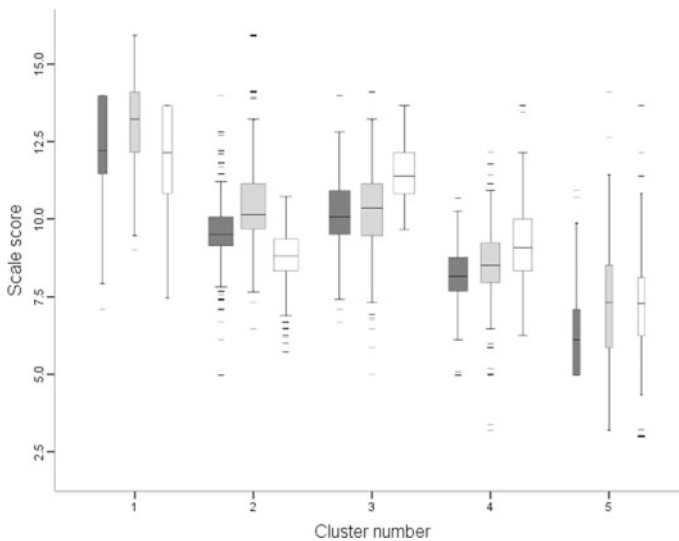


Fig. 4.13 Distributions of motivation variables by cluster for Australia, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.13 Descriptive statistics by cluster for Australia, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	11.5	30.5	26.0	23.3	8.7
Mean plausible value	570.5 ^a	517.5 ^b	519.8 ^b	471.7 ^c	446.0 ^d
Female students in cluster (%) [*]	37.8	49.5	44.9	60.6	66.2
Students spending >45 min per week on homework (%)	45.9	39.0	48.6	37.1	28.0
Mean home educational resources scale score	11.6 ^a	11.1 ^c	11.3 ^b	11.0 ^c	10.7 ^d

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $tl > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 286.415, p < 0.001$)

small difference in enjoyment in favor of cluster 3. There were also significant differences between clusters with respect to gender, with boys overrepresented in high-performing clusters 1 and 3, and girls overrepresented in low-performing clusters 4 and 5. The mean levels of home educational resources also differed significantly across clusters; clusters with higher motivation scores also had more resources. There were also differences in time spent on homework: more students reported spending >45 min per week on homework in the high-performing clusters.

4.3.2 Canada-Ontario

Five clusters were extracted from the Ontario sample (Fig. 4.14). Cluster 1 includes students with high scores on all three variables. Clusters 2 and 3 present interesting patterns of moderate motivation scores, but cluster 2 students placed a high value on mathematics and cluster 3 students had high confidence in their abilities. Clusters 4 and 5 contain students with lower score distributions; although cluster 5 students had lower scores than cluster 4, they equally valued mathematics.

Clusters 1 and 3 do not differ significantly in mean achievement (Table 4.14); although enjoyment and value distributions are much lower for cluster 3 students, their confidence scores were only slightly lower than those of cluster 1 students. Cluster 2 has a much lower mean achievement despite rather high value scores. Cluster 5 has significantly lower mean achievement despite having moderate value scores, rather similar to those of clusters 3 and 4. Gender composition differed across clusters. Girls were underrepresented in cluster 1 which is consistently high for all three motivation variables; meanwhile, girls were overrepresented in the least motivated cluster (cluster 5). Of note is the high-performing cluster 3, where the percentage of boys and girls is similar. Home educational resources differed as

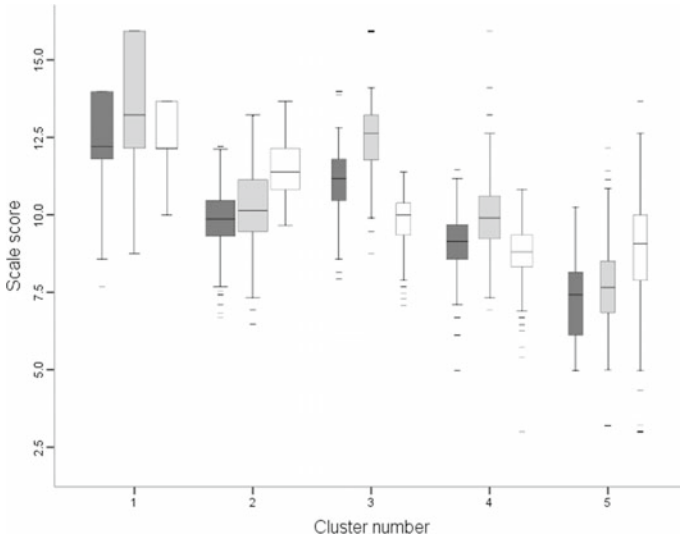


Fig. 4.14 Distributions of motivation variables by cluster for Ontario, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.14 Descriptive statistics by cluster for Ontario, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	21.5	29.0	14.4	21.1	14.0
Mean plausible value	567.9 ^a	514.7 ^b	564.6 ^a	505.0 ^b	462.9 ^c
Female students in cluster (%) [*]	40.4	52.4	49.6	49.7	61.3
Students spending >45 min per week on homework (%)	57.1	65.8	52.3	57.3	58.2
Mean home educational resources scale score	11.7 ^a	11.3 ^b	11.4 ^{a,b}	11.0 ^c	10.9 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 69.135, p < 0.001$)

expected, cluster 1 had higher levels of home resources than cluster 2, while cluster 4 and 5 scored low on this variable. Differences in the percentages of students engaging in homework across the clusters were not large.

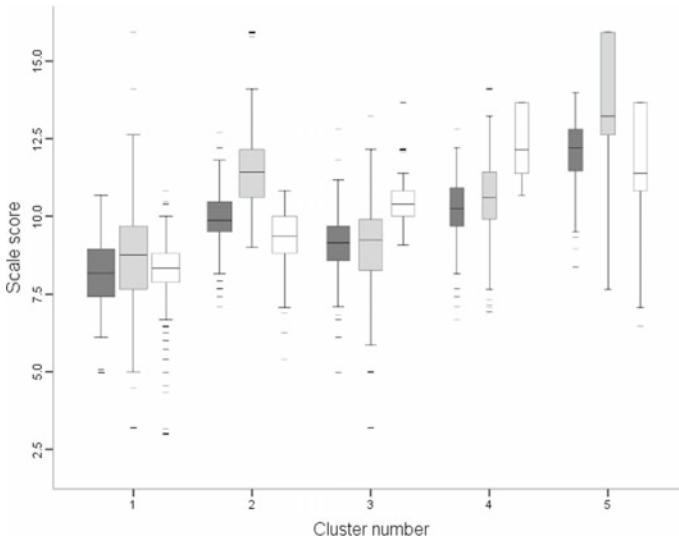


Fig. 4.15 Distributions of motivation variables by cluster for Quebec, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

4.3.3 Canada-Quebec

Five clusters were extracted from the Quebec sample (Fig. 4.15). Distributions of motivation variables reveal that cluster 5 consisted of students who scored highly for all three variables, while cluster 1 consisted of students who scored low on all three variables. The middle three clusters presented some inconsistent profiles with more moderate ranges: cluster 2 was distinguished by relatively high confidence and cluster 4 by higher value scores, while cluster 3 is similar to cluster 4 but contains a lower overall distribution of scores.

Mean student performance was higher in clusters 5 and 2 with high confidence scores (Table 4.15). Students in cluster 4 placed a high value on mathematics, but their mean achievement was lower than those in cluster 5. Gender distribution varied across clusters; there were more girls than boys in the lower motivation clusters. In general, clusters with a higher motivation scores were associated with more home educational resources, and lower percentages of students who reported spending >45 min per week on homework.

4.3.4 England

Five clusters were extracted from the English sample (Fig. 4.16). Two clusters had consistent profiles: cluster 5 is students who scored highly on all motivation variables and cluster 1 is students who scored low on the motivation variables. However, cluster

Table 4.15 Descriptive statistics by cluster for Quebec, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	26.8	24.0	22.3	13.3	13.6
Mean plausible value	510.7 ^a	573.4 ^{b,c}	531.4 ^d	559.6 ^b	588.5 ^c
Female students in cluster (%)*	56.9	48.5	58.6	44.4	48.6
Students spending >45 min per week on homework (%)	75.4	68.1	81.1	76.3	60.0
Mean home educational resources scale score	10.6 ^c	11.0 ^b	10.9 ^{b,c}	11.1 ^{a,b,c}	11.3 ^a

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 42.725, p < 0.001$)

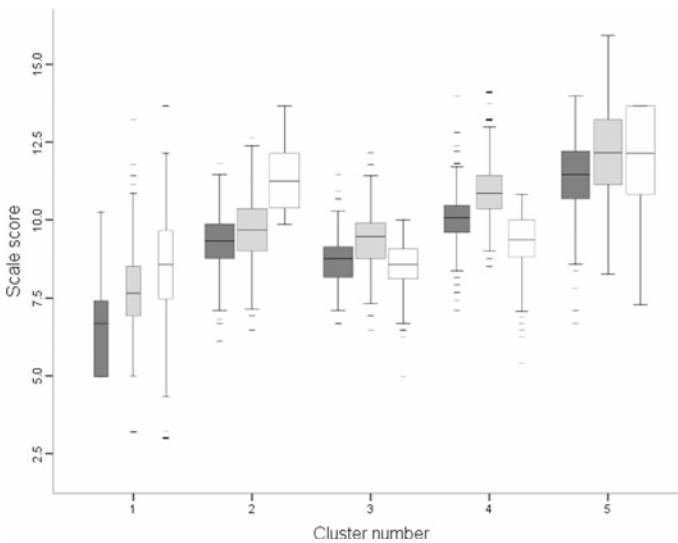


Fig. 4.16 Distributions of motivation variables by cluster for England, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

5 mean student achievement was not significantly higher than that of the cluster 4 students, although cluster 4 had lower distributions for all motivational variables (Table 4.16). Clusters 2 and 3 showed much lower mean achievement despite high endorsement of value in cluster 2. As mentioned, cluster 1 had the lowest mean achievement. Gender composition differed significantly across clusters, with more boys than girls in the highest motivation clusters, and more girls than boys in clusters with lower motivation and achievement. The two highest performing clusters had also

Table 4.16 Descriptive statistics by cluster for England, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	11.2	23.3	23.1	20.3	22.1
Mean plausible value	466.8 ^c	502.8 ^b	501.5 ^b	549.6 ^a	556.7 ^a
Female students in cluster (%)*	66.9	50.0	59.2	47.5	36.3
Students spending >45 min per week on homework (%)	19.1	27.2	22.7	26.6	33.7
Mean home educational resources scale score	10.7 ^c	10.8 ^{b,c}	10.7 ^c	11.0 ^{a,b}	11.2 ^a

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|\text{t}| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 176.879, p < 0.001$)

significantly higher mean home educational resources than the lowest performing clusters. More students reported spending >45 min per week on homework in the high motivation clusters.

4.3.5 Hong Kong

Four clusters were extracted from the Hong Kong sample (Fig. 4.17). Patterns for the motivation variable distributions were quite consistent: for cluster 3, all distributions were high, for cluster 4, all were moderate-to-high, for cluster 2, all were moderate-to-low, and for cluster 1, all were low. Mean achievement differed significantly among clusters and was positively related to the level of the motivation variables (Table 4.17). There were significantly fewer girls than boys in the two highest motivation clusters, and more girls than boys in the other two clusters. The highest performing cluster had a significantly higher score for home educational resources score than the other three clusters; the latter were broadly similar for this SES measure. Student reported time spent on homework did not differ much across clusters.

4.3.6 Hungary

Five samples were extracted from the Hungarian sample (Fig. 4.18). Cluster 1 had high score distributions, and cluster 5 had low score distributions for all motivation variables. Cluster 2 had distinctly high confidence, but low value scores. Clusters 3 and 4 had similar enjoyment and confidence distributions, but differed in their value scores. Clusters 1 and 2 were similar in median confidence, differed for enjoyment and value, but did not significantly differ in mean achievement (Table 4.18). Clusters

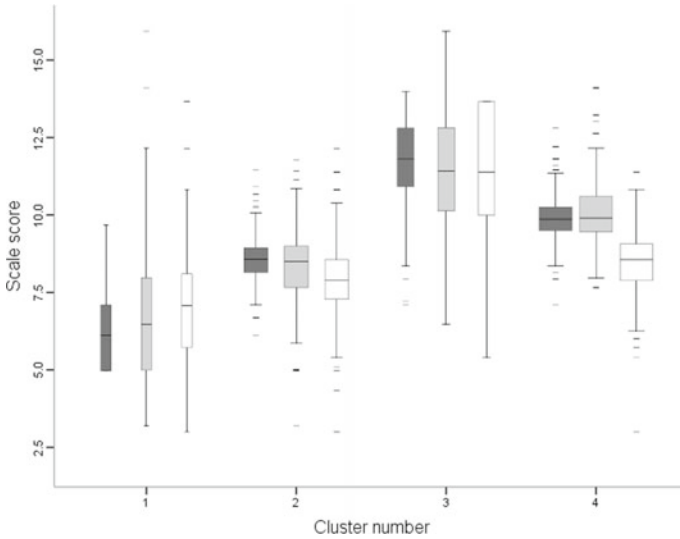


Fig. 4.17 Distributions of motivation variables by cluster for Hong Kong, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.17 Descriptive statistics by cluster for Hong Kong, TIMSS 2015 grade 8

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	12.4	28.6	19.8	39.3
Mean plausible value	546.6 ^d	572.3 ^c	631.9 ^a	606.7 ^b
Female students in cluster (%) [*]	56.8	57.8	32.4	45.2
Students spending >45 min per week on homework (%)	63.0	68.2	60.5	65.6
Mean home educational resources scale score	10.0 ^b	10.0 ^b	10.6 ^a	10.2 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 144.759, p < 0.001$)

3 and 4 had lower but similar mean achievement scores, but differed primarily in their assessment of value. Gender differences were prominent; cluster 1 had the highest motivation and contained more boys than girls, and cluster 5 had the lowest motivation and contained more girls than boys. Clusters 1 and 2 had significantly higher mean scores for home educational resources compared to the other three clusters. Clusters moderate in motivation and achievement reported spending more time on homework than other clusters.

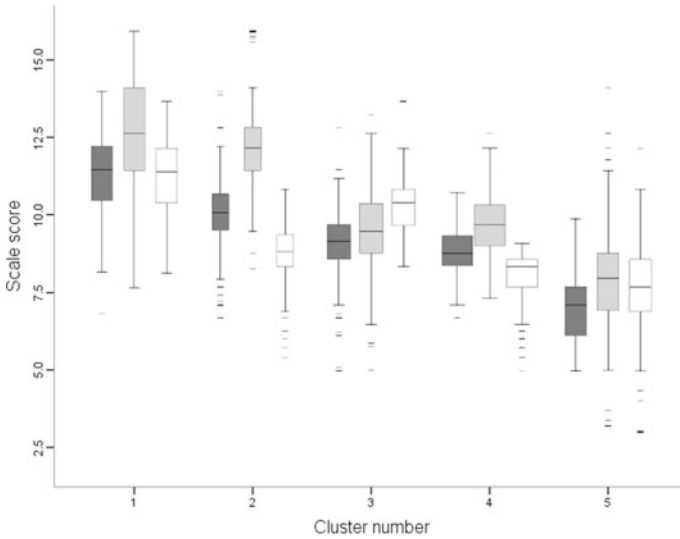


Fig. 4.18 Distributions of motivation variables by cluster for Hungary, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.18 Descriptive statistics by cluster for Hungary, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	18.6	14.6	20.4	27.3	19.0
Mean plausible value	570.6 ^a	570.3 ^a	491.8 ^{b,c}	496.7 ^b	469.7 ^c
Female students in cluster (%) [*]	43.0	50.1	49.6	53.2	56.1
Students spending >45 min per week on homework (%)	48.0	42.2	59.1	51.8	47.9
Mean home educational resources scale score	11.4 ^a	11.3 ^a	10.6 ^b	10.5 ^b	10.5 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 36.268, p < 0.001$)

4.3.7 Iran

Five clusters were found in the Iranian sample (Fig. 4.19). Cluster 5 had consistently high scores on all motivational variables. Cluster 4 had high scores on enjoyment and confidence, but moderate value scores. Cluster 3 had the reverse pattern, with high scores for value, and moderate enjoyment and confidence. The other two clusters had consistent profiles, with moderate-to-low (cluster 2) and low (cluster 1) score distributions for all three variables.

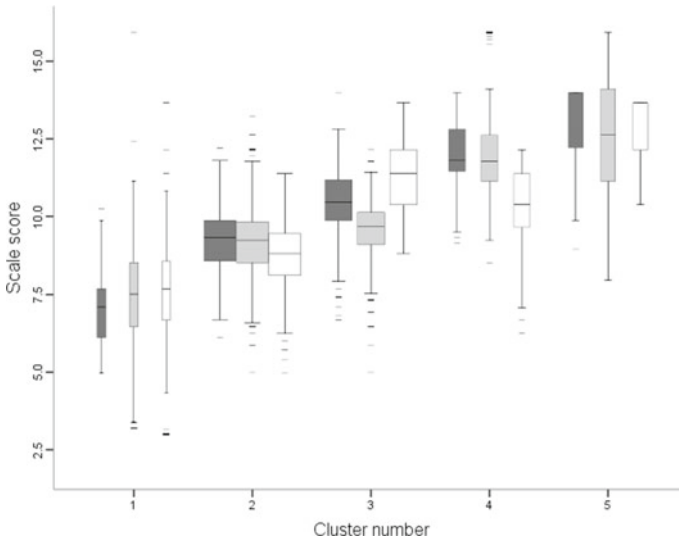


Fig. 4.19 Distributions of motivation variables by cluster for Iran, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Surprisingly, cluster 4 in Iran, with moderate scores for value, had higher mean achievement than the consistently high cluster 5 (Table 4.19). Significantly lower mean achievement was found for clusters 2 and 3, which performed, on average, equally well, despite their differences in value and enjoyment scores. Cluster 1 had the lowest mathematics achievement scores. There were significant differences in the gender composition of clusters; notably, there were equal proportions of boys and girls in the best performing cluster (cluster 4), but more boys in the more motivated and second highest performing cluster (cluster 5). The two clusters with

Table 4.19 Descriptive statistics by cluster for Iran, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	8.6	32.4	27.5	16.6	14.9
Mean plausible value	396.6 ^d	416.9 ^c	418.1 ^c	493.6 ^b	473.2 ^a
Female students in cluster (%)*	54.6	52.4	47.6	50.4	40.1
Students spending >45 min per week on homework (%)	49.7	55.5	59.3	65.4	67.3
Mean home educational resources scale score	9.2 ^{b,c}	9.1 ^b	9.1 ^b	9.9 ^a	9.7 ^{a,c}

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|\text{t}| > 3.29$

*Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 46.370, p < 0.001$)

the higher achievement scores (clusters 4 and 5) had higher average levels of home educational resources, as well as more students reporting spending >45 min per week on homework than the other three groups in Iran.

4.3.8 Japan

Five clusters were derived from the Japanese sample (Fig. 4.20). Most of them had consistent profiles on the three motivation variables: cluster 2 had the lowest median scores, followed by cluster 1, then cluster 3 (which had a rather low score distribution for value), cluster 4, and, finally, cluster 5; the last had high median scores, except for the value variable.

All five clusters differed systematically on average achievement: higher mean achievement was consistent with higher enjoyment and confidence scores, but did not vary consistently with the value scores (Table 4.20). There were large differences in the gender composition of clusters: in the two more motivated clusters there were more boys than girls, while in the two lowest-performing clusters, there were more girls than boys. The level of mean home educational resources did not differ significantly for clusters 4 and 5, but were generally at lower levels in clusters with lower motivation and achievement scores. Students in Japan also reported spending more time doing homework in the clusters with lower motivation scores.

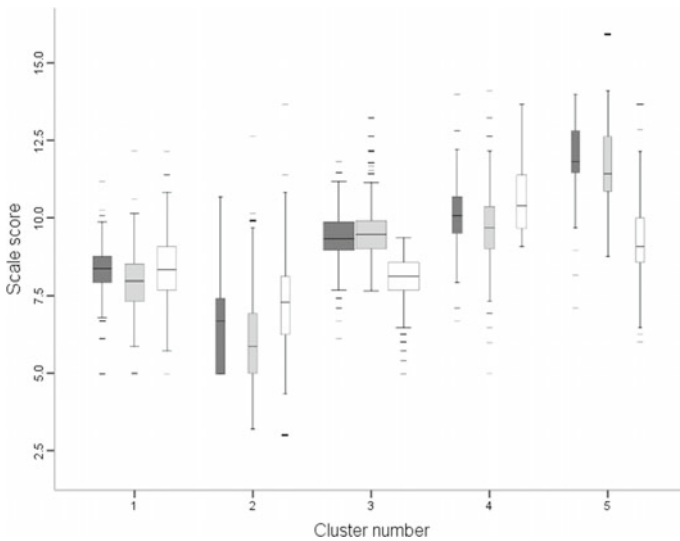


Fig. 4.20 Distributions of motivation variables by cluster for Japan, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.20 Descriptive statistics by cluster for Japan, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	24.6	11.7	40.2	13.0	10.5
Mean plausible value	556.4 ^d	522.3 ^c	596.3 ^c	619.8 ^b	651.5 ^a
Female students in cluster (%)*	59.5	60.0	51.4	39.3	34.7
Students spending >45 min per week on homework (%)	30.8	34.9	25.1	27.1	19.8
Mean home educational resources scale score	10.8 ^{b,c}	10.6 ^b	11.0 ^c	11.3 ^a	11.3 ^a

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $tl > 3.29$

*Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 137.618, p < 0.001$)

4.3.9 Norway

Six clusters were extracted from the Norwegian grade eight sample (Fig. 4.21). Cluster 6 had high score distributions on all three motivation variables, cluster 5 had high confidence scores but otherwise moderate scores on the other two motivation variables, cluster 4 had a high value score, clusters 1 and 3 both had moderate-to-low scores but their mean value scores differed, and cluster 2 had generally low scores for all variables.

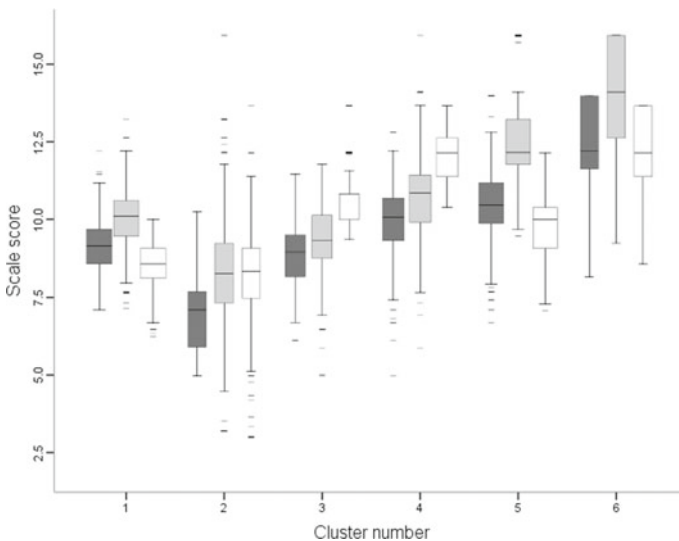


Fig. 4.21 Distributions of motivation variables by cluster for Norway, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.21 Descriptive statistics by cluster for Norway, TIMSS 2015 grade 8

Cluster characteristics	Cluster					
	1	2	3	4	5	6
Size (% of total number of students)	19.9	14.3	15.6	16.9	19.2	14.0
Mean plausible value	478.9 ^b	443.8 ^d	462.3 ^c	481.3 ^b	523.7 ^a	535.1 ^a
Female students in cluster (%) [*]	47.5	57.6	59.4	47.3	48.0	42.9
Students spending >45 min per week on homework (%)	56.5	63.0	62.1	61.1	51.8	50.7
Mean home educational resources scale score	11.3 ^{a,b}	11.0 ^a	11.1 ^{a,b}	11.4 ^b	11.6 ^c	11.8 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 61.050, p < 0.001$)

Differences in mean achievement were significant in most pairwise comparisons, where clusters with higher motivation scores also had higher achievement scores (Table 4.21). Cluster 1 was an exception, as it (a) had higher mean achievement scores than cluster 3, despite cluster 3 having higher value scores; and (b) did not differ significantly from cluster 4, despite the latter having higher motivation scores, particularly for value. Females were again underrepresented in the higher motivation clusters and overrepresented in the lower motivation clusters. Mean home educational resources also differed significantly across the clusters; in particular, the two higher performing clusters reported higher mean levels for home educational resources than the other clusters for this SES measure. In clusters with higher motivation scores, fewer students spent >45 min per week on homework.

4.3.10 Singapore

Six clusters were extracted from the Singapore sample (Fig. 4.22). Cluster 1 had high score distributions on all three motivation variables and for cluster 2 there were also high, but the value distribution scores were lower than in cluster 1. Cluster 3 had a consistently moderate profile, and cluster 6 was similarly moderate, but the value distribution scores were again lower than in cluster 3. Cluster 4 had consistently low score distributions for all three variables, and for cluster 5 these were even lower.

Mean achievement comparisons in Singapore indicate some interesting patterns. When we compared clusters 1 and 2, we found that the latter cluster performed better on average in TIMSS despite the lower distribution scores for enjoyment and especially for value (Table 4.22). Similarly, students in cluster 6 performed better than those in cluster 3, despite having lower value scores. The clusters with low and very low patterns of motivation scores also scored low on mean achievement. The only important difference in gender composition was noted for cluster 1, which contained significantly more boys than girls. Mean home educational resource levels

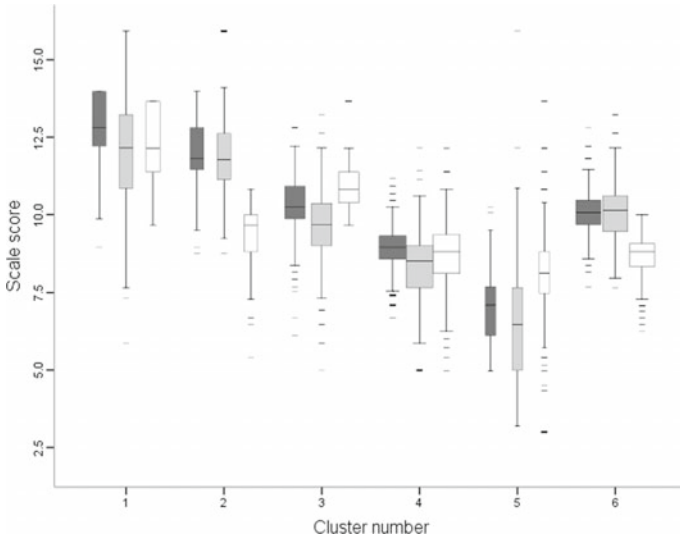


Fig. 4.22 Distributions of motivation variables by cluster for Singapore, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.22 Descriptive statistics by cluster for Singapore, TIMSS 2015 grade 8

Cluster characteristics	Cluster					
	1	2	3	4	5	6
Size (% of total number of students)	12.6	12.3	18.4	24.6	9.5	22.7
Mean plausible value	652.3 ^a	666.6 ^b	616.8 ^c	597.0 ^d	559.0 ^e	632.6 ^f
Female students in cluster (%) [*]	38.8	48.6	49.1	52.6	52.0	48.5
Students spending >45 min per week on homework (%)	77.5	75.3	79.2	76.9	69.3	75.5
Mean home educational resources scale score	10.6 ^a	10.6 ^{a,b}	10.4 ^{a,b}	10.2 ^c	10.0 ^c	10.4 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $ltl > 3.29$

^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(5) = 41.695, p < 0.001$)

did not differ much, but some differences were significant, especially between the high motivation clusters and the low motivation clusters. Compared to the other clusters, a lower percentage of students in the least motivated cluster (cluster 5) reported spending >45 min per week on homework.

4.3.11 Slovenia

Six clusters were extracted from the Slovenian sample (Fig. 4.23). These included a high motivation cluster (cluster 1), a moderate cluster with slightly higher confidence scores (cluster 2), a moderate cluster with high value scores (cluster 3), an inconsistent profile of low confidence and value, with very low enjoyment (cluster 4), a cluster that scored consistently low for all three variables (cluster 5), and another cluster that scored low on value and very low for enjoyment and confidence (cluster 6).

Mean achievement comparisons in Slovenia revealed that the consistently high profile cluster 1 had the highest mean achievement, followed by cluster 2, which had moderate motivation scores (Table 4.23). Interestingly, clusters 3, 4, and 5 did not differ significantly, despite the large differences among these clusters, particularly in the enjoyment and value score distributions. The very low confidence and enjoyment cluster 6 had the lowest mean achievement score. The clusters differed notably in their gender composition; cluster 1 contained more boys than girls, and cluster 6 contained more girls than boys. The mean home educational resources scale score differed among clusters in most comparisons with higher scores in the clusters with the higher motivation distributions. Higher percentages of students reported spending >45 min per week on homework in the clusters with the lower motivation and achievement scores.

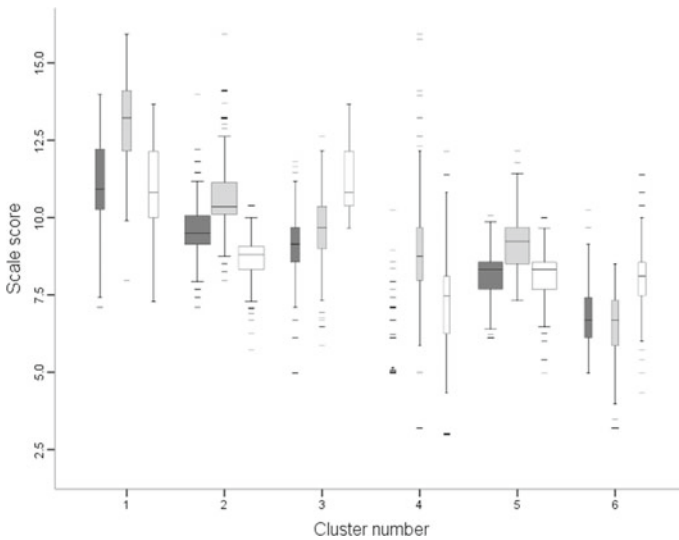


Fig. 4.23 Distributions of motivation variables by cluster for Slovenia, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.23 Descriptive statistics by cluster for Slovenia, TIMSS 2015 grade 8

Cluster characteristics	Cluster					
	1	2	3	4	5	6
Size (% of total number of students)	11.2	31.3	11.2	7.4	29.9	9.0
Mean plausible value	577.0 ^a	538.8 ^b	503.5 ^c	492.0 ^c	496.4 ^c	465.1 ^d
Female students in cluster (%)*	42.5	47.8	44.2	41.7	50.5	61.1
Students spending >45 min per week on homework (%)	41.6	61.0	68.8	70.5	71.7	77.6
Mean home educational resources scale score	11.3 ^a	10.9 ^b	10.9 ^b	10.7 ^{b,c}	10.6 ^c	10.4 ^c

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 42.523, p < 0.001$)

4.3.12 USA

Five clusters were extracted from the USA sample (Fig. 4.24). Cluster 4 had consistently high scores for all three motivational variables, and cluster 5 was similar, but with slightly lower scores on all variables. There were two clusters with moderate scores for confidence and enjoyment, which differed in their value scores; cluster 3 had a low score distribution for the value variable, while cluster 2 had a higher score

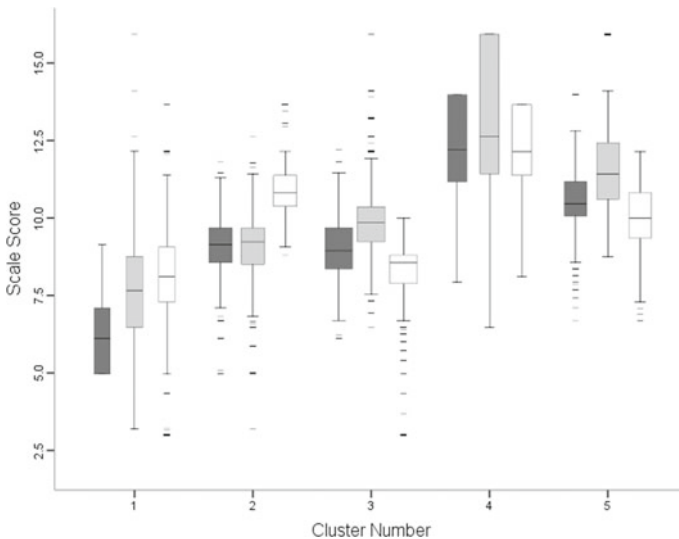


Fig. 4.24 Distributions of motivation variables by cluster for USA, TIMSS 2015 grade 8
Notes Dark gray = enjoyment, light gray = confidence, white = value. Box width represents relative cluster size

Table 4.24 Descriptive statistics by cluster for USA, TIMSS 2015 grade 8

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	14.9	21.3	25.3	17.8	20.7
Mean plausible value	479.5 ^a	494.1 ^b	509.7 ^c	558.9 ^d	553.1 ^d
Female students in cluster (%)*	57.9	54.0	47.6	44.0	51.1
Students spending >45 min per week on homework (%)	56.8	57.9	52.7	50.9	52.0
Mean home educational resources scale score	10.7 ^a	10.8 ^a	10.7 ^a	11.2 ^b	11.1 ^b

Notes Different superscripts (a, b, c, etc.) indicate significantly different mean PV or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$

*Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 81.820, p < 0.001$)

distribution for the value variable. Cluster 1 had consistently low scores for all three motivational variables.

Clusters 4 and 5 did not differ significantly in their mean achievement scores despite the notable differences in the enjoyment and value distribution scores (Table 4.24). Cluster 3 had a significantly lower achievement on average, and cluster 2 had even lower mean achievement, despite high endorsement of the value of mathematics. Cluster 1 had the lowest average achievement score. There were significant differences in the gender composition of clusters; again there were more boys than girls in the highest motivation score cluster and more girls than boys in the lowest motivation score clusters. Levels of mean home educational resources were significantly higher for the two highest performing clusters. Reported time spent on the homework differed only slightly between clusters in the USA.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 5

Cluster Analysis Findings Over 20 Years of TIMSS



Abstract Student motivation clusters were identified by analyzing IEA's Trends in International Mathematics and Science Study (TIMSS) mathematics data from 1995, 2007, and 2015 at both grade four and grade eight in 12 jurisdictions. Deeper examination of the distributions of student motivation variables within each cluster revealed that high confidence scores, followed by high enjoyment scores, were generally associated with better student achievement in mathematics. Students' value for the subject was dissociated from the other two motivation variables in at least one cluster in nearly all samples. Statistical tests indicated that mean achievement, gender composition, and mean home resources scores differed across clusters in systematic ways at both grades and across all three TIMSS administrations. Twenty years of TIMSS data reveal interesting patterns in the number of clusters, the relationships among the three motivational variables and their association with achievement, and concerning the gender composition of distinct motivational clusters. Clusters with similar mixes of motivational variables seem to produce consistent and almost universal effects across jurisdictions.

Keywords Background variables · Confidence · Demographic variables · Enjoyment · Interaction of motivation variables · Motivation clusters · Trends in International Mathematics and Science Study (TIMSS) · Value

5.1 Summary of Cluster Analysis Results for Grade Four Samples

We used measures of enjoyment of and confidence in mathematics as our input variables for the cluster analysis of grade four samples and examined solutions comprising three to five clusters.

5.1.1 *The TIMSS 1995 Administration*

At grade four, in six of the TIMSS 1995 jurisdictions that we examined, four clusters were extracted, and five clusters were extracted from five jurisdictions¹ when enjoyment and confidence in mathematics variables were used as our input. Appendix B gives a comprehensive summary of the distributions of the motivation variables by cluster and jurisdiction, while Appendix C provides achievement and demographic characteristics by cluster and jurisdiction.

Score distributions for enjoyment and confidence often overlapped within a single cluster. When similar within a cluster, higher scale values were associated with higher mean achievement scores. In every sample, there was at least one cluster where the distributions of enjoyment and confidence did not overlap. Comparison of such inconsistent clusters with the others, enabled us to examine the association with mean cluster achievement. The picture was not conclusive. First, there were jurisdictions (such as England, Hungary, Iran, and the USA) where mean achievement was not significantly different across all clusters. For example, in Hungary three clusters out of four did not significantly differ in mean achievement. Second, there were cases where neither enjoyment nor confidence varied systematically with mean achievement across all clusters (e.g., in Hungary and Singapore). Then, there were samples where the distribution of enjoyment seemed to be more aligned with mean achievement (such as Australia, England, and Iran), and samples where the distribution of confidence seemed to be more aligned with mean achievement (as seen in Ontario, Quebec, Japan, Slovenia, and the USA).

The gender composition across clusters was also variable. Differences across clusters were statistically significant, but usually not large, and there was sometimes no clear pattern in the gender composition of different clusters. Boys were overrepresented in the high motivation and achievement clusters and underrepresented in the low motivation and achievement clusters in the Japanese and the Hong Kong samples, and this was also true in Slovenia, but the differences were less extreme. In England and Iran there were small, but statistically significant, differences in gender composition; girls were overrepresented in high motivation and achievement clusters and underrepresented in low motivation and achievement clusters. We noted non-significant gender differences in Ontario (sample size for Ontario in 1995 was rather small), and inconsistent patterns in the remainder of the samples. As regards homework, clusters with more motivated and higher-achieving students reporting undertaking more than one hour of homework daily were found in Japan and Australia; in the other samples, such differences were small, or the patterns were inconsistent.

¹Norway participated in the 1995 administration, but not with grade four and grade eight students (for further details, see Chap. 3). We therefore excluded the 1995 data from Norway from our analyses, and conducted cluster analysis for TIMSS 1995 grade four for only 11 jurisdictions.

5.1.2 *The TIMSS 2007 Administration*

In nine out of the 12 jurisdictions, our cluster analysis produced five distinct clusters, while in the remaining three countries (Australia, Iran, and Norway) we derived four clusters (see Appendix B for a comprehensive summary of the distributions of the motivation variables by cluster and jurisdiction, and Appendix C for achievement and demographic characteristics by cluster and jurisdiction).

At least one cluster in every sample could be described as consistent, with similar score distributions for confidence and enjoyment in mathematics. When the two distributions reflected high motivation scores, the mean cluster achievement was high; when the two distributions reflected low motivation scores, mean achievement was low.

There were many instances of inconsistent profiles where scores on confidence and enjoyment differed within a cluster (namely, where clusters contained high scores on one of the motivation variables and low scores on the other variable). It was evident in all cases that cluster mean achievement was more closely associated with confidence than with enjoyment scores. Even if patterns looked slightly different between countries, it was clear that motivational profiles with high confidence in mathematics were adaptive for performance. High enjoyment of mathematics appeared to be less important, and was not always linked to high achievement, unless it was coupled with a high level of confidence. While evident in all samples, this was particularly salient in Ontario, Quebec, Hong Kong, Norway, Singapore, and the USA.

In terms of gender composition, the differences among clusters were statistically significant in all samples, typically more boys tended to be present in the highly motivated clusters and/or more girls tended to be present in the lower motivation clusters. In actual percentages, there were generally only small differences in the gender composition of the clusters within each country. The most common finding was that there were fewer girls in the most motivated clusters. We noted the largest differences in gender composition in the Hong Kong sample. Small discrepancies in the proportions of boys and girls within a cluster were found in Iran, Norway, and the USA. In the majority of the jurisdictions, students in the clusters that had low motivation scores reported spending more time on homework, although in Iran, England, and Singapore the opposite was generally true.

5.1.3 *The TIMSS 2015 Administration*

We extracted five clusters in eight of the jurisdictions, and four clusters in each of the remaining samples (see Chap. 4). Examination of the distributions of motivation variables revealed that enjoyment and confidence had similar distributions within clusters, although there was at least one cluster in each sample where scores for enjoyment and confidence differed. In most countries, the clusters scoring most highly for motivation and achievement were relatively small in size, with the

exception of Singapore. Boys were the majority of students in the high motivation clusters in most jurisdictions, with the exception of Iran.

When comparing confidence and enjoyment, confidence was generally more closely associated with achievement compared to enjoyment. In most samples, there was evidence of at least one pair of clusters where, despite similar levels of confidence, higher enjoyment was linked to lower achievement; Hungary and Norway present examples of such patterns.

There were gender differences in cluster composition across all countries, except Iran. Typically, boys were overrepresented in clusters with higher motivation scores and girls overrepresented in clusters with lower motivation scores. Differences in home resources for learning were statistically significant, but also rather small. Cluster mean scores for home resources for learning were usually positively associated with average achievement, except in Norway and Hungary, where the pattern was not systematic.

5.2 Summary of Cluster Analysis Results for Grade Eight Samples

At grade eight, students responded to questions that were used to create three motivation scales: enjoyment of, confidence in, and value for mathematics. We performed cluster analyses on the scores on these three scales, and examined solutions of three to six clusters in all samples.

5.2.1 *The TIMSS 1995 Administration*

From the 11 samples in the study from the TIMSS 1995 administration (Norway data were not analyzed in 1995, see Chap. 3), we generally extracted five clusters, except for the Japanese and the Hong Kong samples, where four clusters were identified. Appendix B presents a comprehensive summary of the distributions of the motivation variables by cluster and jurisdiction, while Appendix C conveys achievement and demographic characteristics by cluster and jurisdiction.

In general, patterns of cluster differences were more conclusive in grade eight than in grade four. When the distributions of the motivation variables were consistently high on the response scale for a cluster, mean achievement was also high; when the opposite was true, mean achievement was low. England was a unique case where there were apparently very few significant differences in achievement between clusters. In all 11 jurisdictions, some clusters showed inconsistent distributions of motivation variables; in these clusters the distribution of value for mathematics was typically higher than the confidence in and enjoyment of mathematics distributions; however, such motivation profiles did not correspond with higher mean achievement for the

cluster. The distributions for confidence and enjoyment were very similar in most clusters. When they differed in a cluster, levels of confidence were more strongly aligned with mean achievement than levels of enjoyment in about half of the samples (namely, Ontario, Quebec, Japan, Slovenia, and the USA), but this was not the case for the remainder of the samples.

The gender composition of clusters differed significantly across clusters in most of the samples. Typically, there were more boys than girls in the high motivation clusters and more girls than boys in the low motivation clusters. There were no significant differences in the gender composition of clusters in Hungary, Iran, and Slovenia. There were higher percentages of parents with parental education above secondary level in the high motivation and achievement clusters, but the differences were small for most of the samples. Homework was coded as the percentage of students engaging in more than one hour of homework daily. Homework engagement across clusters was not systematically associated at the cluster level with motivation and achievement levels. Across multiple clusters, high motivation clusters reported a high percentage of students engaging in more than one hour of homework daily, while low motivation clusters reported lower percentages of students engaging in more than one hour of homework daily.

5.2.2 The TIMSS 2007 Administration

Following cluster analyses, six clusters were extracted in nine jurisdictions, five clusters in two, and the Slovenian sample provided four clusters (see Appendix B for a comprehensive summary of the distributions of the motivation variables by cluster and jurisdiction, and Appendix C for achievement and demographic characteristics by cluster and jurisdiction).

In each of the 12 samples, there were clusters where the distributions of the motivation variables were consistent (i.e., a group of students with high, moderate, or low ratings for all three motivation variables). As expected, the mean achievement for a cluster was positively associated with the average standing for the motivation variables.

However, there were also inconsistent mixtures across all jurisdictions that suggest additional information about the relative significance of confidence in, enjoyment of, or value for mathematics to achievement. Value for mathematics was often dissociated from the other two variables in at least one, and usually in two clusters per sample. There were often two clusters with low or moderate scores for confidence and enjoyment, but much higher ratings for value, implying that there were groups of students extrinsically motivated, while being intrinsically less motivated. Those clusters implied that value was related to achievement only when it was consistent with confidence and enjoyment (i.e., averaging equally high or low levels). The only exception to this was demonstrated by cluster 3 in the sample from the USA, where the high value and lower confidence and enjoyment distributions were associated with moderately high mean achievement scores compared to other clusters.

Typically, mean achievement was positively associated with the relative levels of the other two variables, even when value scores were dissociated.

Confidence and enjoyment scores were in general more consistent within clusters, although there were cases where their distributions did not overlap extensively. In those cases, it was the level of confidence, rather than enjoyment, that matched mean cluster achievement levels.

Gender composition differed significantly across clusters in all samples, except in Iran, where the gender by cluster test of independence was non-significant. In general, there were more boys than girls in high motivation and high achievement clusters and more girls than boys in the clusters with low motivation and achievement scores. In some cases, such as England, the differences in percentages were large in many clusters; in others, larger discrepancies in gender composition appeared only within the high achieving groups (e.g., in Australia and Hungary), or in low achieving groups (e.g., in the USA). Yet in other jurisdictions, discrepancies were overall minor (e.g., in Norway, Singapore, and Slovenia).

Differences in parental education by cluster were identified in all samples. The outcome variable was the percentage of parents who had completed post-secondary education or above. For students in lower performing clusters the reported percentage of parents who had completed post-secondary education or above was smaller than that for students in higher performing clusters. Regarding homework engagement, in half of the jurisdictions there was no systematic pattern detected; in some samples, highly motivated clusters reported more homework engagement (e.g., in Australia and Iran), and in other samples, low motivation clusters reported spending more time on homework (e.g., in Ontario and the USA).

5.2.3 The TIMSS 2015 Administration

We extracted five clusters in eight jurisdictions, four clusters from Hong Kong, and six clusters in three jurisdictions (see Chap. 4). Our examination of the distributions of motivation variables within the clusters indicated that there were both consistent and inconsistent motivation profiles in a sample. Hong Kong was the only country with only consistent cluster profiles. Across all countries, mean achievement was higher for clusters where students endorsed motivation statements strongly. This was particularly noticeable in profiles with consistent motivational scores: in most cases, clusters of students with the highest enjoyment of, confidence in, and value for mathematics distributions also had the highest mean achievement; the opposite was true for the clusters with the lowest motivation scores.

When profiles were inconsistent, interesting achievement differences were found in many cases. With the exception of Hong Kong, where all clusters had consistent score profiles, all other jurisdictions presented evidence that high endorsement of the value of mathematics was not associated with high performance (Quebec, Hungary, Norway, and Singapore are key examples). In at least one cluster across all samples, the value for mathematics score distribution was often dissociated from

the other two motivation variables. Enjoyment and confidence had similar score distributions within most clusters. Confidence was the variable most closely related to mean achievement. In the few cases where the distribution of enjoyment differed from that of confidence within a cluster, enjoyment seemed to be the variable that was less closely aligned with mean achievement.

There was a consistent pattern of gender composition across samples: there were significantly more boys than girls in the clusters with high motivation score distributions. Iran was an exception, where the second highest in motivation, but top performing cluster had a balanced gender composition. In contrast, clusters with low motivation scores contained more girls than boys across all samples.

The home educational resources variable, which is used as a proxy for SES, usually displayed only small within-country differences, but was consistently associated with motivation. Clusters with more motivated students had higher mean home educational resources scores than clusters with lower motivation distributions. There were no consistent results for the homework variable across jurisdictions: in some samples, clusters with more motivated and higher achieving students reported doing more homework than their counterparts (e.g., in Australia, England, and Iran). In other samples, the higher motivation clusters reported spending less time on homework than their peers (e.g., in Quebec, Japan, Norway, and Slovenia), and for some countries there were no systematic patterns (e.g., Ontario and Singapore).

5.3 Twenty-Year Patterns in TIMSS by Country and Grade

We undertook cluster analyses for 12 jurisdictions, two grades and across three cycles of TIMSS. While comparisons should be made with caution because technical characteristics were not constant across the various administrations of TIMSS (see Chap. 3 for details), we were able to identify some interesting patterns in our analyses. Motivation variables were measured using different sets of items (particularly in 1995, when only two or three items were used for confidence and enjoyment), and scoring for the motivation variables was done by averaging item means in 1995 and 2007, and by using latent methods in 2015. Sampling procedures also differed in the TIMSS 1995 administration, hence sample sizes for this study were small for some jurisdictions (Ontario in particular). Moreover, the analysis was conducted at the level of student groups with particular profiles; patterns and differences are examined at the cluster level and inferences should not be drawn at the level of the individual student.

Cluster analysis is an exploratory procedure and the decision for determining the number of clusters in this study was based on judgmental criteria, as described in Chap. 3. Across all three administrations, four or five clusters were derived from each grade four sample, and four to six clusters from each grade eight sample (see Table 5.1). We considered solutions with three clusters, however in those cases only consistent profiles of motivation variables were observed. Analyses that resulted in more than six clusters led to clusters that were too small in size and to clusters with

Table 5.1 Number of clusters extracted from cluster analysis by grade, jurisdiction, and administration

Jurisdiction	Number of clusters extracted					
	Grade 4			Grade 8		
	1995	2007	2015	1995	2007	2015
Australia	5	4	4	5	6	5
Ontario	4	5	5	5	5	5
Quebec	4	5	5	5	6	5
England	4	5	4	5	6	5
Hong Kong	5	5	5	4	6	4
Hungary	4	5	5	5	6	5
Iran	4	4	5	5	6	5
Japan	5	5	4	4	5	5
Norway	–	4	5	–	6	6
Singapore	5	5	4	5	6	6
Slovenia	5	5	5	5	4	6
USA	4	5	5	5	6	5
Average number of clusters	4.45	4.75	4.67	4.82	5.67	5.17

very similar profiles with no additional interpretative value. On average, we extracted fewer clusters at grade four than at grade eight across all administrations. This is an artifact of the analysis, because at grade eight one extra input variable, value, was used, and solutions of six clusters were examined. In 1995, fewer clusters were obtained than in the subsequent administrations for both grades; however warranted claims about trends in cluster numbers cannot be made considering the limitations we have already discussed in the previous paragraph.

Confidence and enjoyment variables had overlapping score distributions in many clusters. However, in every cluster analysis, there was at least one cluster for which the two variables had distributions with different score ranges. By examining the two distributions across all clusters, it was possible to identify which one most closely aligned with mean cluster achievement (Table 5.2). In most cases, the confidence score distribution was more closely associated with achievement (i.e., clusters with higher confidence scores tended to have higher mean achievement in mathematics). In a few cases, this did not occur when confidence and enjoyment had overlapping distributions in all clusters (e.g., in Japan for TIMSS 2015, grade eight, and indeed in several jurisdictions for TIMSS 1995, grade eight; see Table 5.2). This overlap made it difficult to discern which of the two variables was more clearly associated with achievement. Enjoyment of mathematics at the cluster level was more closely aligned with mean achievement than confidence in some cases in the TIMSS 1995 grade four samples (e.g., in Iran).

Table 5.2 Relative importance of confidence and enjoyment variables on achievement across administrations by grade and jurisdiction

Jurisdiction	Grade 4			Grade 8		
	1995	2007	2015	1995	2007	2015
Australia	✗	✓	✓	✗	✓	✓
Ontario	✓	✓	✓	✓	✓	✓
Quebec	✓	✓	✓	✓	✓	✓
England	✗	✓	✓	✗	✓	✗
Hong Kong	✗	✓	✓	✗	✗	✗
Hungary	✗	✓	✓	✗	✓	✓
Iran	✗	✓	✓	✗	✓	✓
Japan	✓	✓	✗	✓	✓	✗
Norway	–	✓	✓	–	✓	✓
Singapore	✗	✓	✓	✗	✓	✓
Slovenia	✓	✓	✓	✓	✓	✓
USA	✓	✓	✓	✓	✓	✓

Notes ✓ confidence was more important than enjoyment in the association with mean achievement; ✗ confidence was equally or less important than enjoyment in the association with mean achievement. In 1995, Norway’s grade 4 and grade 8 students did not participate

Value for mathematics was a contextual variable measured only at grade eight. Scores distributions for value were often dissociated from the confidence and enjoyment variables. We commonly observed that ratings for value were higher than ratings for the other two variables in at least one cluster in most samples. Within each jurisdiction, an examination of how the levels of the three distributions varied across clusters and the change in the cluster mean achievement enabled us to evaluate the alignment of the value variable with mean achievement. There was some evidence (i.e., in at least one cluster) that scores for value did not covary in the expected direction with mean achievement in every sample (see Table 5.3). For example, a cluster would have a distribution with high value scores when the mean achievement for the cluster was low. In these very common cases, the confidence and enjoyment variables would also tend to score low, and to be more strongly aligned with mean achievement.

For every sample, we tested whether cluster membership was independent of gender. In most cases, this test was statistically significant (see Table 5.4) suggesting that the proportion of boys and girls in each cluster was not independent from the cluster categorization. Inspection of percentages in each cluster revealed that boys tended to be overrepresented in the clusters with high motivation and achievement scores, and/or underrepresented in clusters with lower motivation and achievement scores. Sometimes, there were unbalanced proportions of boys and girls in the intermediate clusters without a systematic pattern, but the highest or the lowest performing clusters (and often both) usually reflected the general trend. The exception

Table 5.3 Alignment of the valuing mathematics distributions with achievement across administrations by jurisdiction (grade eight only)

Jurisdiction	TIMSS cycle		
	1995	2007	2015
Australia	✘	✘	✘
Ontario	✘	✘	✘
Quebec	✘	✘	✘
England	✘	✘	✘
Hong Kong	✘	✘	✓
Hungary	✘	✘	✘
Iran	✘	✘	✘
Japan	✘	✘	✘
Norway	–	✘	✘
Singapore	✘	✘	✘
Slovenia	✘	✘	✘
USA	✘	✘	✘

Notes ✘ value distributions not associated with mean achievement across all clusters; ✓ value distributions aligned with mean achievement in all clusters. In 1995, Norway's grade 4 and grade 8 students did not participate

Table 5.4 Gender composition of clusters across administrations by grade and jurisdiction

Jurisdiction	Grade 4			Grade 8		
	1995	2007	2015	1995	2007	2015
Australia	✓	✓	✓	✓	✓	✓
Ontario	✘	✓	✓	✓	✓	✓
Quebec	✓	✓	✓	✓	✓	✓
England	✓ ^a	✓	✓	✓	✓	✓
Hong Kong	✓	✓	✓	✓	✓	✓
Hungary	✓	✓	✓	✘	✓	✓
Iran	✓ ^a	✓ ^b	✘	✘	✘	✓ ^b
Japan	✓	✓	✓	✓	✓	✓
Norway	–	✓	✓	–	✓	✓
Singapore	✓	✓	✓	✓	✓	✓
Slovenia	✓	✓	✓	✘	✓	✓
USA	✓	✓	✓	✓	✓	✓

Notes ✓ the chi-square test (gender × cluster) was significant at 0.05; ✘ the chi-square test (gender × cluster) was not significant at 0.05. In 1995, Norway's grade 4 and grade 8 students did not participate. ^aMore girls were present in the high motivation and achievement clusters and more boys were present in the low motivation and achievement clusters. ^bThere were approximately equal percentages of boys and girls in the top-performing group

was Iran: the tests of independence for the Iranian sample were either statistically non-significant, or if significant, girls were overrepresented or at least equally represented in the cluster with the highest motivation and achievement scores.

Overall the cluster evaluation revealed consistent findings at both grades, and across TIMSS administrations and jurisdictions. Examination of the distributions of motivation variables in each cluster reveals that clusters with high confidence scores were most strongly linked to better mathematics achievement, followed by those that scored highly for enjoyment. These two variables had overlapping score distributions within most clusters. Scores for value for the subject did not overlap with the other two motivation variables in at least one cluster in almost all samples. Statistical tests showed that mean achievement, gender composition, and mean home resources scores differed significantly across clusters in systematic ways at both grades and across all three TIMSS administrations. There were a few exceptions. In the TIMSS 1995 administration samples, it was not easy to discern differences between confidence and enjoyment; it is likely that changes in levels of student motivation have occurred after a decade, but the scores used in the TIMSS 1995 analysis may also be less reliable, because these two motivation variables were measured using just two or three items. Iran provides quite distinct gender comparisons, and plausible explanations may include cultural differences. Chapter 6 provides critical discussion of the findings, along with connections to the research literature and implications for research and practice.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 6

Insights from Motivational Profiles in TIMSS Mathematics



Abstract A person-centered cluster analysis approach to the study of motivation in IEA's Trends in International Mathematics and Science Study (TIMSS) mathematics has revealed interesting profiles of students across key motivational constructs. Between four and six different clusters were extracted from each sample analyzed. Unsurprisingly, some clusters had consistent motivation scores, but in almost every jurisdiction, there were clusters of students with inconsistent score distributions between the contributing motivational constructs. The clusters were systematically different on various external variables, such as mean mathematics achievement, gender composition, and the level of home resources available to students. The study also presents a novel way of looking at the relative importance of enjoyment of, confidence in, and value for mathematics, and the association of these motivation variables with achievement and other demographic characteristics at the cluster level. When motivation scores were mixed rather than consistent, there was a uniform achievement advantage enjoyed by the groups of students who had higher scores for confidence in mathematics over enjoyment of, or value for mathematics. This approach revealed that gender and socioeconomic background are not independent of cluster membership. Typically, clusters with high confidence values were comprised of more boys than girls, and students from better resourced homes. The findings can be linked to relevant literature on motivation in mathematics. Educational efforts to develop student motivation need to take into account differential student profiles and prioritize techniques that target skill and competence in mathematics.

Keywords Cluster analysis · Educational achievement · Family characteristics · Mathematics competence · Mathematics motivation · Student characteristics · Student motivation profiles

6.1 Examining the Role of Motivation in Educational Achievement

From the early endeavors of IEA to study educational achievement cross-culturally (such as the Six Subject Survey, and the First International Mathematics Study; see <https://www.iea.nl/other-iea-studies>), a broad objective was to understand the

relationships between inputs and outputs in education more fully (Wagemaker 2014). With insights from this research, innovations and reforms could then be designed to assist in the improvement of educational systems. Over the past 60 years, ILSA programs have developed, expanded, and increased to encompass multiple school subjects and skills, a range of populations, diverse item formats and modes of administration, advanced technical and analytic methods, and detailed contextual and demographic measures. Hypotheses and research questions can be addressed through primary and secondary analyses of ILSA data. Moreover, successive administrations of each ILSA program facilitate the examination of stability or change in the measured characteristics.

In our study of student motivation profiles, we analyzed grade four and grade eight data from the TIMSS 1995, 2007, and 2015 administrations for 12 jurisdictions, focusing specifically on the mathematics assessments. We explored three dimensions of student motivation: enjoyment of mathematics (enjoyment), self-confidence in mathematics ability (confidence), and perceived value of mathematics (value, which was assessed at grade eight only). The number of items in the student background questionnaires that tap motivational, self-concept, and affective constructs related to mathematics have more than doubled over time, from 10 items in TIMSS 1995 to 27 items in TIMSS 2015.

Motivational variables are considered important predictors of achievement, and are even discussed as curricular goals by themselves (Hooper et al. 2013). Research on the relationship between motivation and achievement is abundant in the fields of educational psychology and mathematics learning. It has been examined extensively within various theoretical frameworks, such as self-determination theory, expectancy-value theory, self-efficacy, self-concept, and achievement goal theory. It has been addressed with correlational and experimental data, data from ILSA programs, and from TIMSS in particular. Reviews and meta-analytic studies (Hattie 2009; Lee and Shute 2010) have shown that motivational characteristics are associated with achievement. Similar research endeavors have additionally tried to compare those predictors; self-efficacy and confidence have been found to correlate more strongly with achievement than other variables (Richardson et al. 2012; Stankov 2013). Nonetheless, the strength of the relationship of these motivational and affective variables to achievement tends to be weak ($r < 0.30$). This suggests that while there is a general positive association between achievement and confidence, and interest and value, it may not be universal. It may be that, in looking at the combinations of the motivational constructs, some insight might be shed on factors that limit the association. The traditional correlational approach would suggest that all motivation variables correlate with achievement, and, if modeled together, their relative importance would vary and that they might interact; but the person-centered approach has revealed that groups of students, often large in size, have very distinctive profiles that would remain undetected in a variable-centered analysis.

6.2 Clusters of Students Using Motivation Variables: A Person-Centered Approach

Using a person-centered approach implemented via two-step clustering is a novel approach to the problem in the context of TIMSS motivational characteristics. Instead of looking at linear relationships between each motivation variable and achievement, as is common in the literature, we aimed to identify meaningful clusters of students; clusters with distinct profiles based on varying combinations of enjoyment, confidence, and value scores. These constructs are positively inter-correlated, but they are not identical, hence students need not score consistently on all three. Clustering is an empirical method for grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other clusters. As such, we used the two (enjoyment and confidence at grade four) or three (enjoyment, confidence, and value at grade eight) motivational scores provided by student self-reports to group those with similar response patterns, independent of their background variables or their mathematics performance. From each sample within 12 jurisdictions, three administrations, and two grades, we extracted between four and six clusters.

We found two types of patterns in the results. The most obvious pattern, especially in light of the intercorrelation of the variables, is consistent. As expected, there were clusters with consistently strong endorsement for both (in grade four) or all three (in grade eight) variables, clusters with consistently moderate scores, and clusters with consistently low scores for the motivation variables. Given the positive relationship reported in the literature between motivational variables and educational performance, the mean achievement in these clusters was related to the level of motivation. As such, these clusters conform to the results expected from the variable-centered type of analysis.

A more interesting pattern was revealed by the inconsistent clusters: groups of students who were inconsistent in their motivation variable responses. If a student is inconsistent (e.g., scoring high on one variable, but medium or low on other motivation variables), the natural question is whether endorsement of one motivational construct over the others is associated with greater learning success. Thus, rather than presuming that each contributing motivational variable works the same way for all participants, this clustering approach shows that the motivational variables work in quite different ways for subsections of the student population, and potentially explains why the overall effect of the variable toward achievement is relatively modest. The most common inconsistent profile was that of high value for mathematics and lower enjoyment and confidence for the subject. Inconsistent standing on confidence and enjoyment was also detected in multiple samples, although to a lesser extent; the two variables had distributions that usually aligned. Within the inconsistent clusters, mean achievement was higher when the score distribution for confidence, and often enjoyment, was high, while scores for value had less of an effect.

We also examined the clusters for differences in student characteristics and family background, (i.e., gender composition, home resources or parental education, and homework engagement). With respect to student gender, there were more girls than boys in clusters with low motivation, while there were more boys than girls in clusters with high motivation, or clusters that had high confidence but lower scores on other variables. With respect to family background, clusters with higher motivation scores and greater achievement had more home resources. The effect of more privileged homes was also reflected by the relative strength of confidence; students from more privileged homes tended to have greater confidence in mathematics. These findings validate the person-centered approach as theoretically worthwhile, and highlight the usefulness of identifying clusters with well-defined characteristics.

6.3 Motivation Clusters and Achievement

The role of the value for mathematics variable was a consistent finding across jurisdictions and administrations at grade eight. Value has to do with recognizing the importance, worth, or usefulness of something. This is inherently extrinsic, since the value of mathematics lies in how contemporary society uses mathematical skills to solve technical and technological problems. Because of that, the value of mathematics often lies in its utility and its contribution to future educational and employment opportunities for school students. Hence, as well as lying within the individual, value in mathematics also lies external to the individual in how it is viewed and treated by society. Recognition of the value of mathematics is meant to inspire students to persevere with learning it and prioritizing it in their schoolwork. But it is fundamentally an external motivator, and may even be accompanied with dislike for the subject (Ryan and Deci 2000). External motivators have been found to be less strongly associated with achievement (Eklöf 2007; Lee and Stankov 2018; Marsh et al. 2006).

In consistent clusters, the value score distribution would overlap with the confidence and enjoyment distributions. In inconsistent clusters, value would diverge; value would be typically higher than the confidence and enjoyment variables, implying that there are groups of students who value mathematics and consider the subject important in their life, but do not feel capable or intrinsically motivated. The cluster analysis identified such distinct groups of students. However, their motivational profile was usually less academically successful, since ascribing a higher value was not associated with higher achievement. When dissociated from enjoyment and confidence, a high value for mathematics could not compensate for low interest and self-competence, and was associated with comparatively low levels of performance. A recent multi-level analysis by Lee and Stankov (2018) with multiple predictors of achievement from TIMSS and PISA has also shown that value has a negative effect size, much smaller than self-beliefs, affect, and other motivation variables. A similar suggestion was made by Ratelle et al. (2007) after identifying motivational profiles of students within a SDT framework; students with an autonomous profile had similar performance but higher perseverance than students with a combined autonomous and extrinsically-oriented motivational profile.

The comparison of enjoyment of and confidence in mathematics was also systematic across countries, administrations, and the two grades. Confidence in personal ability to perform learning tasks has been found to be a stronger predictor for achievement (e.g., Stankov 2013; Lee and Stankov 2018; Marsh et al. 2006). Enjoyment and positive affect towards mathematics are desirable attitudes influencing engagement with the subject and fostering learning. Students with greater interest in a domain tend to have greater prior knowledge and learn more from instruction (Murphy and Alexander 2002). Score distributions for the two variables were largely overlapping in most of the clusters (with or without consistent scores on the value variable). This pattern is consistent with research among older students (Marsh et al. 2006); strong covariation has been found between affect that intrinsically motivates students and belief in their own mathematical competence. However, in the current study, when they were not consistent within a cluster, it was confidence that was more closely associated with achievement.

An implication from the current analysis is that a positive attitude towards mathematics is adaptive (i.e., associated with greater achievement) if coupled with confidence and efficacy beliefs. Nonetheless, this is not the naïve notion of believing in confidence itself or in oneself; rather, this is an example of justified confidence that is associated with actual performance. The students with stronger endorsement of confidence rightly believed they could do the mathematics in the TIMSS tests: they achieved higher scores than their peers who prioritized value or enjoyment, but lacked strong beliefs in their capabilities. Thus, it would seem that mathematics teaching should seek to develop justified confidence in doing mathematics tasks. As students gain competence in the domain of mathematics it seems plausible that they will develop confidence in their capabilities. Trying to give students a sense of confidence, independent of real capability, is unlikely to be effective (see Pajares 2008). Perhaps the challenge for education involves moving classrooms from the practice of teachers making students interested in mathematics or knowing its value, to one in which teachers focus on helping students become competent. It may be sensible with novices to invoke situational interest, but the goal is to lead them to intrinsic interest as a consequence of greater competence, expertise, and knowledge (Murphy and Alexander 2002).

6.4 Motivation Clusters, and Student and Family Characteristics

Consistent gender differences were found between clusters at grades four and eight, across all three administrations and in nearly all samples. It should be noted that we did not compare the mean achievement by gender in this analysis, but the composition of clusters, which also happened to differ in mean achievement. The gender composition differences were typically observed in the extreme clusters, namely the high versus low motivation clusters. There were more boys than girls

in clusters with high motivation students, while there were more girls than boys in clusters with low scores on motivation variables. It is generally reported that boys report greater confidence than girls in mathematics, while in terms of interest the difference is equivocal (Ganley and Lubienski 2016). Hence, the current results from high achieving clusters, whether with consistently high motivation scores or in mixed clusters with high confidence, seem to reflect the advantage boys have around confidence, rather than enjoyment.

The Iranian sample was an exception, in that gender differences were either non-significant, or, when significant, the trend favored girls. It is worth mentioning that, in Iran, students attend single-sex schools, which may reduce sex differences in school achievement. Although Marsh et al. (2013) did not include Iran, a non-Arabic-speaking nation, in their study of four Arab countries, they found comparable results; gender differences on motivation variables and achievement in TIMSS favored girls, in contrast to four Anglo countries where the trend favored boys. Marsh et al. (2013) suggested that in Arab cultures girls have higher achievement because they put more time and effort toward schoolwork compared to boys. Recent studies comparing a large number of countries have found that in more gender equal and more economically developed countries, differences between males and females increased along more traditional lines with respect to (a) occupational and educational preferences (Falk and Hermle 2018), and (b) relative academic strengths and pursuit of science, technology, engineering, and medicine (STEM) degrees (Stoet and Geary 2018). A plausible explanation is that in less gender-equal countries, pressures surrounding the quality of life promote the engagement of females with STEM subjects (Stoet and Geary 2018). In contrast, economic development and gender equality in other countries may allow differential, albeit gender-specific, preferences by males and females to be manifested (Falk and Hermle 2018).

Indicators of socioeconomic background are routinely included in studies of the determinants of educational achievement. Measures such as parental education level, occupation, and income, as well as resources available to students at home, have been used as proxies for SES. Items measuring such characteristics have been included in background questionnaires in TIMSS administered to the students or the parents from the first cycle of the program. When used individually, such indicators are not necessarily stronger than self-belief measures (Lee and Stankov 2018). In the TIMSS 2015 administration, relevant indicators were combined to generate a scale score for home educational resources (grade eight) and home resources for learning (grade four). A multidimensional composite is more likely to result in stronger relationships to achievement (Van Ewijk and Slegers 2010). Mean comparisons revealed statistically significant differences in home resources across clusters in nearly all samples.

Motivation levels, and achievement levels were not independent of the socioeconomic resources measures. Higher resources scores were found in clusters consisting of students high on all motivation variables, or in inconsistent clusters with high confidence. The pattern was evident even though differences in the cluster resources scores within jurisdictions were smaller than between jurisdictions. The important message here is that home resources were positively associated with

competence and motivation in general, as well as achievement. Parents in higher SES homes are more educated themselves, so they are better able to assist with early school learning, model and socialize academic success, and afford extra tuition and other educational resources. These social determinants of achievement are well-established (Coleman 1966) and it would be unreasonable to expect teachers or individual students to take responsibility for such factors.

The differences in home resources across clusters in grade four samples, while statistically significant, appeared to be smaller and not always systematic compared to the differences in grade eight samples. However, it was evident that socioeconomic disparities are present from the elementary school years (see, e.g., Duncan et al. 2011). School policies and teaching should probably prioritize methods that result in faster development of learning for students from lower resourced homes. For example, widespread use of techniques such as spaced practice, interleaving, retrieval practice, elaboration, concrete examples, and dual coding (Weinstein et al. 2018), frequent reviews, asking questions, and providing models and scaffolding (Rosenshine 2010) is likely to reduce knowledge deficits associated with reduced home resources. Policies that aim to reduce gaps associated with socioeconomic factors are warranted on multiple grounds. Social programs and policies designed to ensure that all children are raised in homes with sufficient socioeconomic resources is a responsibility that goes well beyond that of teaching, schools, or education.

Results about cluster differences in time invested on homework were mixed. In the TIMSS 2015 administration, engagement in homework was not consistent across countries: more motivated and higher achieving students reported more homework in some countries, less homework in others, or there was no systematic pattern between homework and cluster membership in other samples. Descriptive statistics from the TIMSS 2007 grade eight and TIMSS 1995 grade four administrations gave a similar picture. Systematic patterns could be detected in a few jurisdictions where more motivated student clusters reported doing more homework (e.g., TIMSS 1995 grade eight) or less homework (e.g., TIMSS 2007 grade four) than their peers. Moreover, the percentages of homework engagement within samples suggest that there are large differences between jurisdictions. Policies for homework are highly variable across cultures. Total time spent in school, type of school or class, prevalence of afternoon programs or private tutoring, and parental involvement in students' learning are also factors that affect the amount of time devoted to homework, and such factors are not homogeneous across cultures (Chen and Stevenson 1989; Dettmers et al. 2009). This result seems consistent with the research on the value of homework, which suggests that it is highly variable depending on a number of factors, including age of students and style of work done at home (Marzano and Pickering 2007). Hence, the relationship between motivation and homework engagement remains complex and contextualized, at least in terms of the cluster-level approach employed in this study.

6.5 Methodological Concerns

The cluster analysis was implemented using student self-reports on items measuring motivation in the TIMSS background questionnaires. As described in Chap. 3, the measurement of the enjoyment, confidence, and value variables has gradually evolved from a few items intermixed in a list of items measuring attitudes towards mathematics in TIMSS 1995, to separately-presented, multi-item scales with evidence supporting their factorial structure in TIMSS 2015. Score estimation was also different across administrations, with the use of item averages in TIMSS 1995 and 2007, and latent variable methods in TIMSS 2015. Adopting a “construct-level” approach, the cluster analyses in the TIMSS 1995, 2007, and 2015 samples were conducted with the motivation variables as inputs, despite differences in their operationalization and, as a consequence, in their psychometric properties. With these limitations in mind, the cluster solutions have not been compared across time (i.e., in terms of relative size or mean mathematics achievement). Beyond the number of extracted clusters for each jurisdiction, which varied across administrations, graphical comparison was restricted to cluster differences within a grade/administration/jurisdiction sample. Cohort changes in motivation would be an interesting research question to explore, particularly when important reforms are implemented in a country. Invariant methods for obtaining data and modeling scores would enable such comparisons.

Enjoyment and confidence variables were available in TIMSS grades four and eight, while items measuring value for mathematics were administered only at grade eight. Using different input variables in cluster analysis for the two grades prevented the direct comparison of motivation clusters across grades. This difference meant that the hypothesis of developmental decline in motivation could not be examined in the context of cluster analysis. Since similar items for enjoyment and confidence were administered to grade four and grade eight students, provided measurement invariance holds, mean comparisons could be a fruitful further study to pursue.

The cohort findings reported in this book are grounded in visual examination of boxplots. The visual display of the distributions of the variables by cluster facilitated the interpretations for each cluster. However, we did not attempt to validate the cluster results. Discriminant analysis may be used to validate the clustering method, but this would not prove that the students in a cluster are actually different to those in a different cluster. It would only demonstrate the feasibility of using predictor variables for matching the atheoretical empirical method of clustering. A potentially useful extension of this work would be to design field studies to describe different clusters in more detail with methodologies beyond survey self-reports. Research in the real world of school classrooms might identify whether students who respond differently on the TIMSS questionnaire items differ in some way in their practice, strategy usage, or learning, and so on.

Finally, the promising results generated by these cluster analyses suggest that similar approaches can be employed in future research in a number of ways. For example, instead of the classic cluster analysis employed in this study, which

uses data-based agglomeration techniques, latent class analysis might provide more subtle and sophisticated student groups, provided there is a valid model for generating latent classes (Vermunt and Magidson 2002). This research may provide a theoretical basis for modeling motivational classes in learning mathematics; further studies could seek evidence to validate these profiles. Moreover, it would be potentially useful to include more time points in a hypothesis-driven framework, to establish better warranted claims regarding longitudinal changes in student motivation and related outcomes. For example, motivational profiles in relation to achievement can be compared before—and after certain policy, curriculum, or pedagogical innovations to study the impact of such changes. Incorporating more than one country that implemented similar policies (e.g., the formal introduction of assessment for learning, or standardized testing for school accountability) would allow the study of the effectiveness of policies across culturally or geographically diverse jurisdictions. Comparison of mean scores for the motivation variables across clusters could also be explored. Assuming measurement invariance holds, clusters could be compared within countries, and across countries which differ in average performance (e.g., high achieving versus low achieving), culture (e.g., individualistic versus collectivistic), language (e.g., Indo-European versus Asian), performance trends in TIMSS (e.g., increasing versus stable versus decreasing), and so on. Studies that focus on clusters of students provide interesting and potentially different results from studies conducted at the individual-student level. However, inferences from a group-level analysis, such as the current analyses, should not be deduced for individuals, to avoid ecological fallacies.

6.6 Concluding Remarks

We used an innovative person-centered, multiple-sample approach, which provided informative findings and some surprising results. While the inconsistent profiles among the motivational variables were anticipated, we were surprised as to how consistent the results were across TIMSS administrations, jurisdictions, and grades. Although jurisdictions varied in location, culture, population size, mean performance, and socioeconomic level, the motivational clusters looked similar, both in terms of motivational patterns and demographic characteristics; this was evident despite the differences in how TIMSS measured motivation in 1995, 2007, and 2015. The similarity of the clusters and their patterns across such a diverse collection of jurisdictions, administrations, and grades strengthen the generality of the findings and add new empirical knowledge to the literature on motivation.

Clusters with students high on all motivation variables typically had the highest achievement. It was revealing though that, when motivation variables were inconsistent in their distributions, higher levels of confidence, often aligned with enjoyment, were always associated with higher mean performance; this was not as pronounced with value for mathematics. This combined result suggests strongly that student groups that have a mixture of motivational characteristics can be

identified. Moreover, gender and socioeconomic background are not independent of the motivation clusters. Accordingly, educational efforts to develop motivation need to take into account differential student profiles, and prioritize techniques that target skill and competence in mathematics over other motivational dimensions.

References

- Chen, C., & Stevenson, H. W. (1989). Homework: A cross-cultural examination. *Child Development*, *60*(3), 551–561.
- Coleman, J. (1966). *Equality of educational opportunity*. Washington, DC: National Center for Educational Statistics.
- Detmers, S., Trautwein, U., & Lüdtke, O. (2009). The relationship between homework time and achievement is not universal: Evidence from multilevel analyses in 40 countries. *School Effectiveness and School Improvement*, *20*(4), 375–405.
- Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology*, *47*(5), 1263–1279.
- Eklöf, H. (2007). Self-concept and valuing of mathematics in TIMSS 2003: Scale structure and relation to performance in a Swedish setting. *Scandinavian Journal of Educational Research*, *51*(3), 297–313.
- Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412), eaas9899.
- Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, *47*, 182–193.
- Hattie, J. A. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon, UK: Routledge.
- Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013). TIMSS 2015 context questionnaire framework. In I. V. S. Mullis, & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 61–82). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/timss2015/frameworks.html>.
- Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in K–12 academic performance: An integrative perspective on student learning. *Educational Psychologist*, *45*(3), 185–202.
- Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, *65*, 50–64.
- Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*(4), 311–360.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., ... & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, *105*(1), 108–128.
- Marzano, R. J., & Pickering, D. J. (2007). The case for and against homework. *Educational Leadership*, *64*(6), 74–79.
- Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *The Journal of Experimental Education*, *70*(3), 197–214.
- Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 111–139). New York, NY: Lawrence Erlbaum Associates.

- Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. *Journal of Educational Psychology, 99*(4), 734–746.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387.
- Rosenshine, B. (2010). *Principles of instruction*. Brussels, Belgium: International Academy of Education.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.
- Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences, 55*(7), 727–732.
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581–593.
- Van Ewijk, R., & Slegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134–150.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars, & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.
- Wagemaker, H. (2014). International large-scale assessments: From research to policy. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 11–36). Boca Raton, FL: CRC Press.
- Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive Research: Principles and Implications, 3*(1), 1. Retrieved from <https://doi.org/10.1186/s41235-017-0087-y>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Appendix A

IBM SPSS Code for the Two-Step Cluster Analysis

```
* Random seed.
SORT CASES BY IDSTUD (A).
set rng mc seed 123456789.
compute randvar=rv.uniform (1,1000).
sort cases by randvar.
delete variables randvar.

* Cluster analysis.
TWOSTEP CLUSTER
/CONTINUOUS VARIABLES=Motivation1 Motivation2 Motivation3
/DISTANCE LIKELIHOOD
/NUMCLUSTERS FIXED=X /* Specify number of clusters.
/HANDLENOISE 0
/MEMALLOCATE 64
/CRITERIA INITHRESHOLD(0) MXBRANCH(8) MXLEVEL(3)
/VIEWMODEL DISPLAY=YES EVALUATIONFIELDS=PV1 PV2 PV3 PV4 PV5
ITSEX <other demographics>
/PRINT IC COUNT SUMMARY
/SAVE VARIABLE=Cluster_noX. /* Save cluster membership variable.

* Descriptives
SORT CASES BY Cluster_noX (A).
SPLIT FILE SEPARATE BY Cluster_noX.

DESCRIPTIVES VARIABLES=PV1 PV2 PV3 PV4 PV5
/STATISTICS=MEAN STDDEV MIN MAX.
FREQUENCIES VARIABLES=ITSEX <other demographics>
/ORDER=ANALYSIS.
SPLIT FILE OFF.

* Chi-square test for categorical variables (e.g., gender).
CROSSTABS
```

```
/TABLES = ITSEX BY Cluster_noX  
/FORMAT = AVALUE TABLES  
/STATISTICS = CHISQ  
/CELLS = COUNT  
/COUNT ROUND CELL.
```


Appendix B

TIMSS 1995 and 2007 Boxplots by Cluster for Each Jurisdiction

In Chap. 4, we presented results by jurisdiction for the TIMSS 2015 grade four and grade eight samples in alphabetical order by country. Scale scores were derived after conducting advanced latent variable methodology for the motivation variables, and for the home resources measure available in the TIMSS 2015 datasets (unavailable in previous administrations). In the TIMSS 2007 and 1995 cycles, simpler procedures were used for scoring: averages of items were estimated to generate the motivation variables, and single items were used as proxies of socioeconomic background. In this appendix, we present the distributions of motivation variables by cluster and jurisdiction for the TIMSS 1995 and 2007 administrations. We present summaries of the cluster analysis results for all administrations and grades, along with trend comparisons, in Chap. 5.

Boxplots depict the distribution of scores for the two (for grade four) or the three (for grade eight) motivational variables for each cluster. In all plots in this section, dark gray shading is used to indicate the enjoyment variable, light gray shading is used to indicate the confidence variable, and boxplots representing the value variable are left unshaded. Box widths reflect the relative cluster sizes. Appendix C provides the descriptive statistics by cluster for cluster size, mean achievement, gender composition, and mean home resources for learning.

B.1 Grade Four

See Figs. B.1, B.2, B.3, B.4, B.5, B.6, B.7, B.8, B.9, B.10, B.11, and B.12.

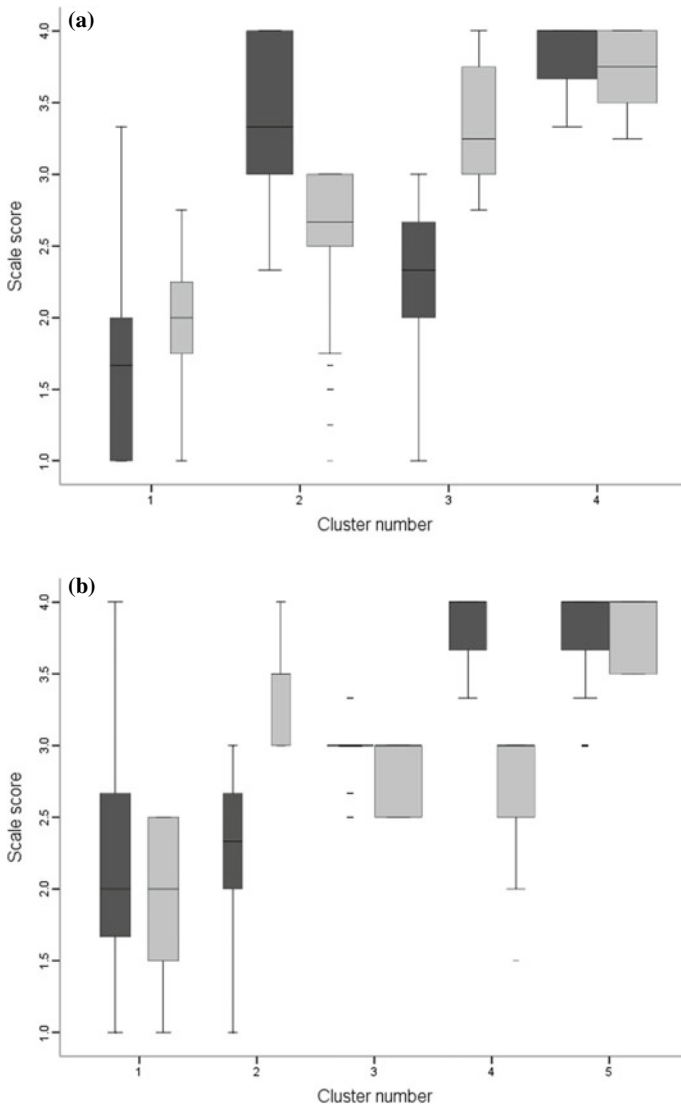


Fig. B.1 Distributions of motivation variables by cluster for Australia at grade 4. (a) 2007, and (b) 1995

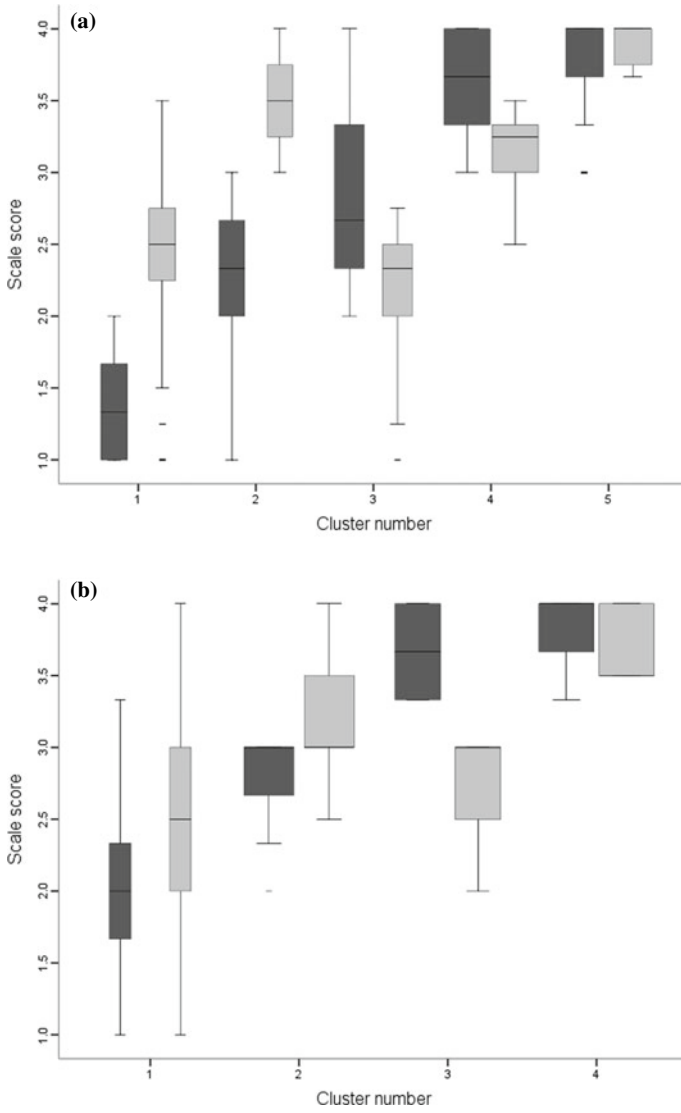


Fig. B.2 Distributions of motivation variables by cluster for Canada-Ontario at grade 4. (a) 2007, and (b) 1995

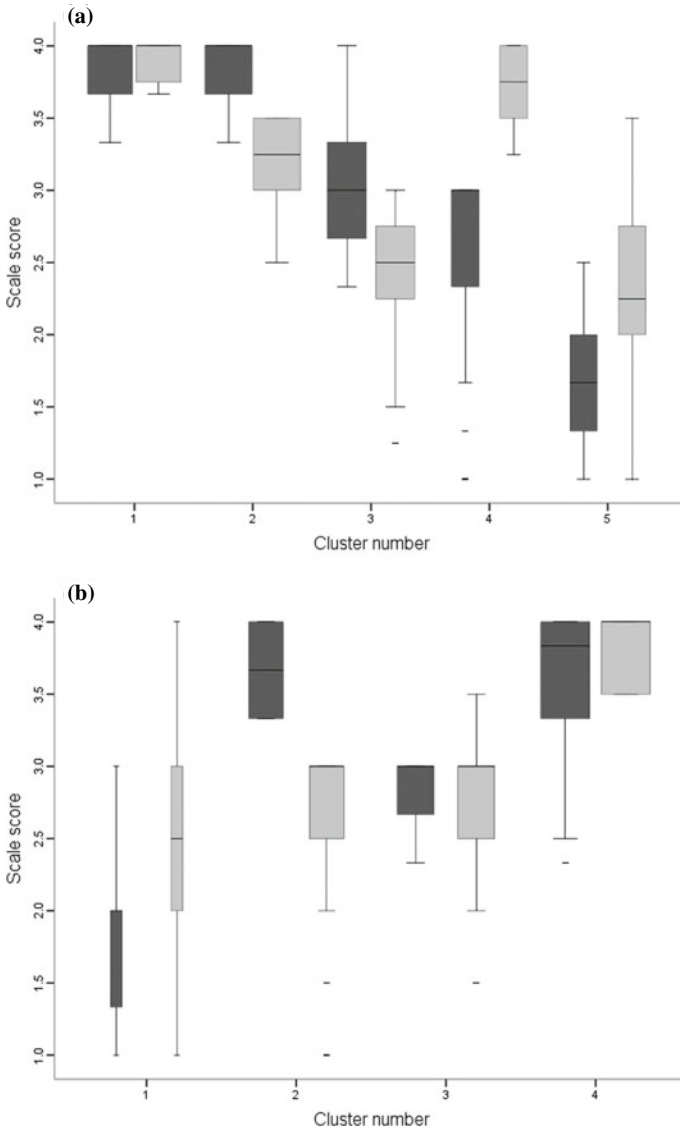


Fig. B.3 Distributions of motivation variables by cluster for Canada-Quebec at grade 4. (a) 2007, and (b) 1995

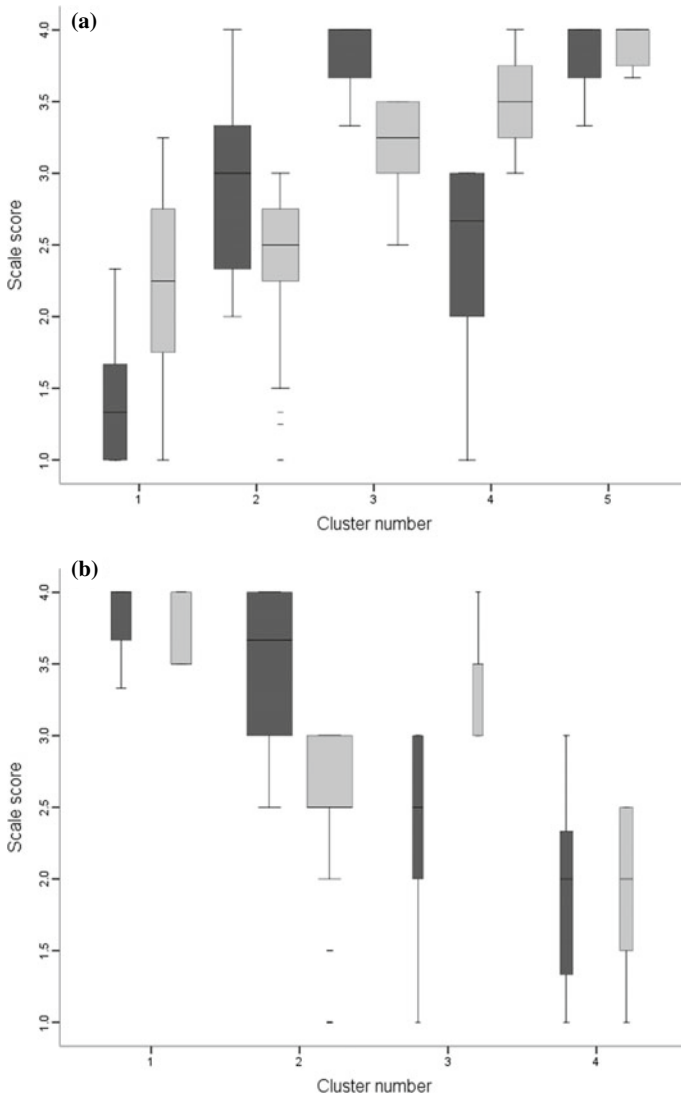


Fig. B.4 Distributions of motivation variables by cluster for England at grade 4. (a) 2007, and (b) 1995

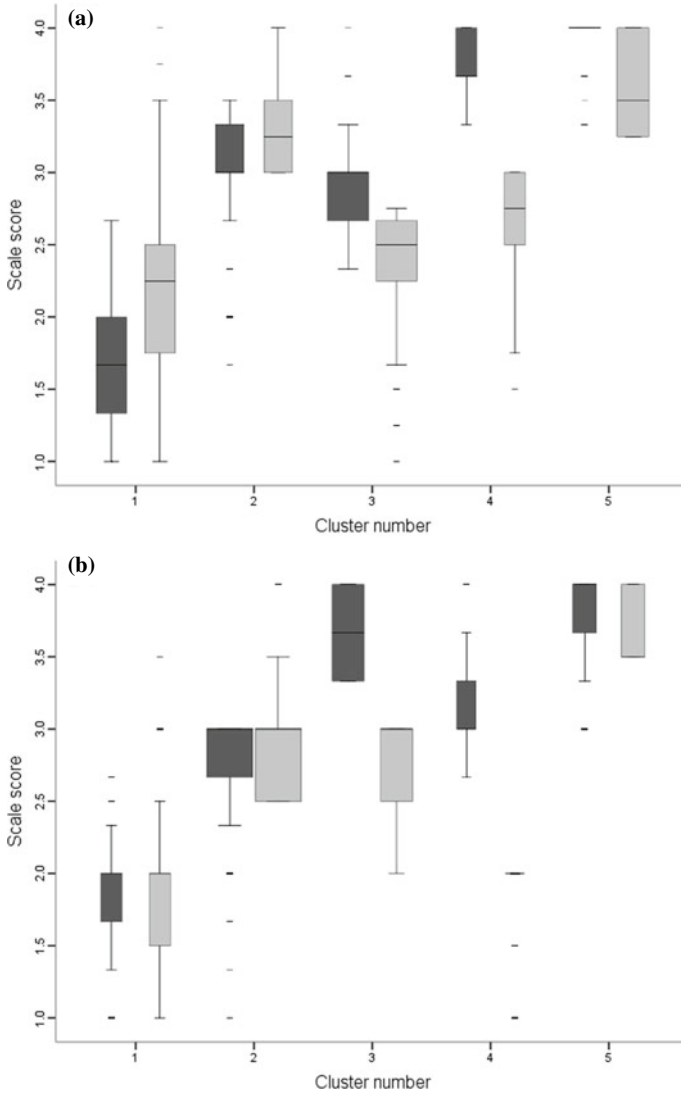


Fig. B.5 Distributions of motivation variables by cluster for Hong Kong at grade 4. (a) 2007, and (b) 1995

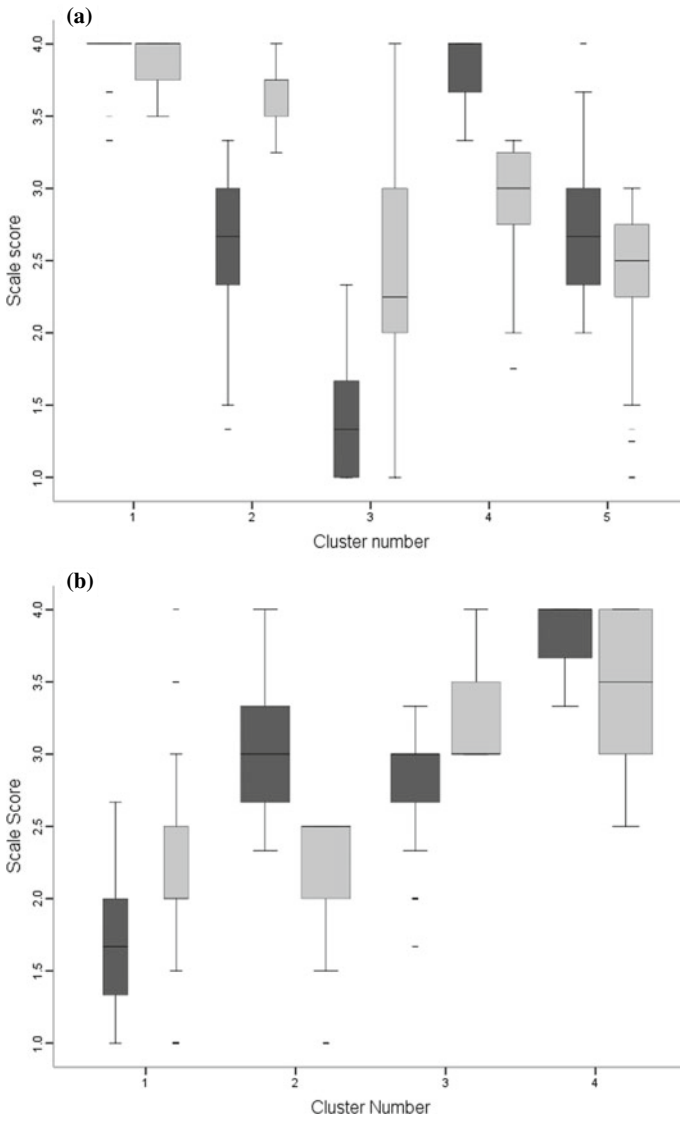


Fig. B.6 Distributions of motivation variables by cluster for Hungary at grade 4. (a) 2007, and (b) 1995

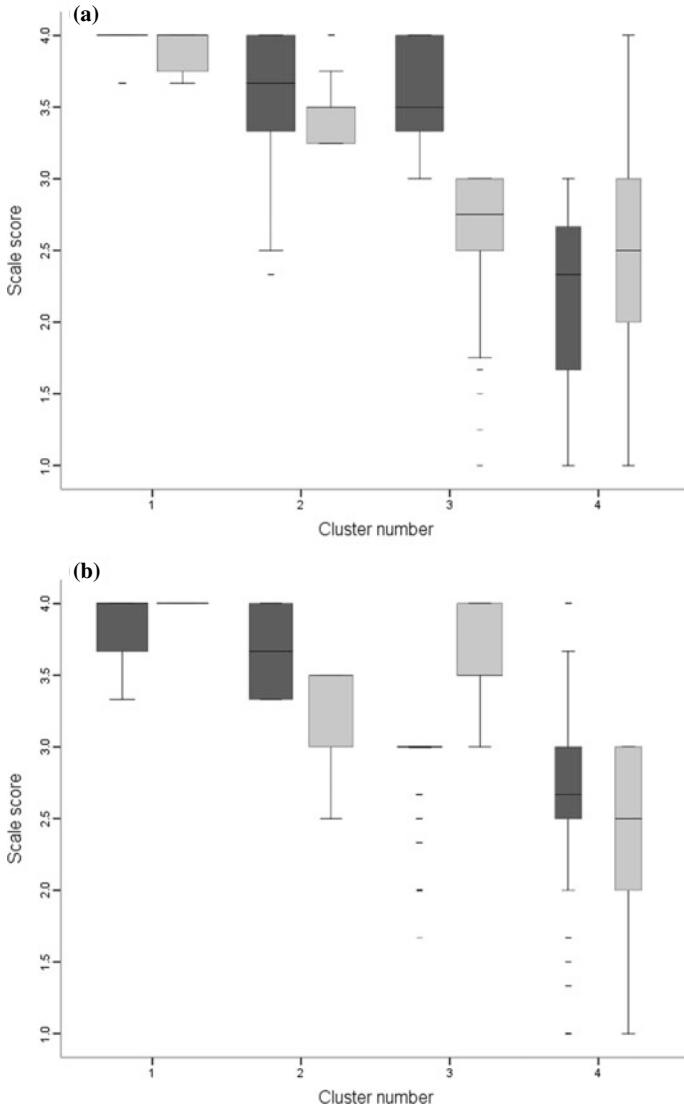


Fig. B.7 Distributions of motivation variables by cluster for Iran at grade 4. **(a)** 2007, and **(b)** 1995

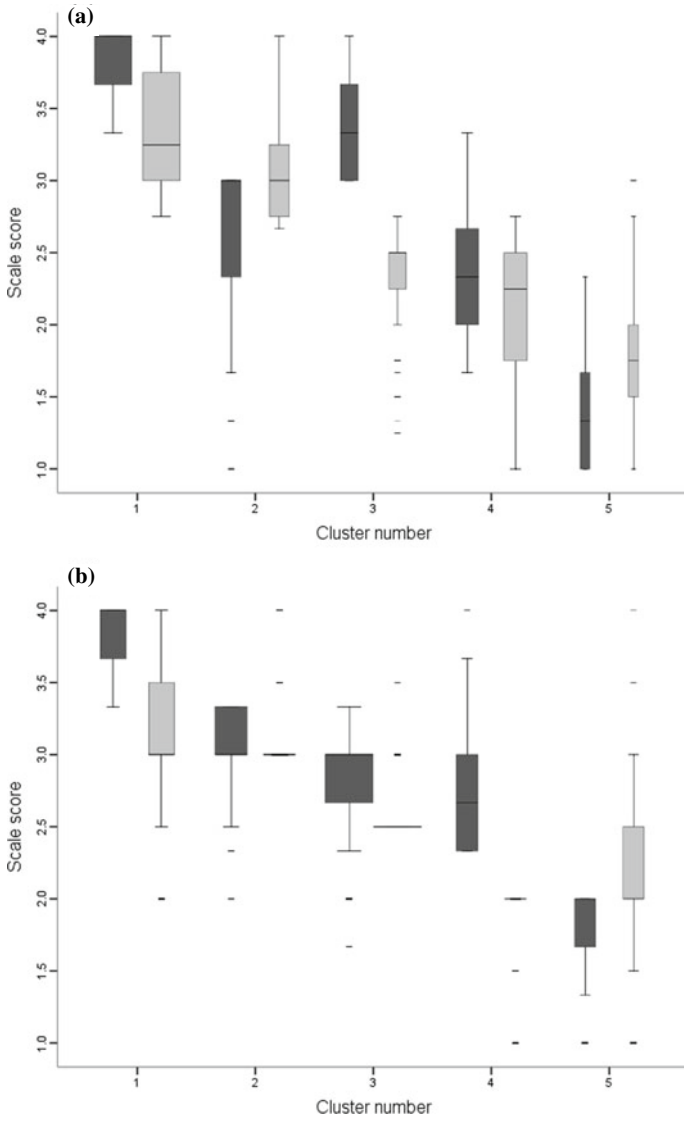


Fig. B.8 Distributions of motivation variables by cluster for Japan at grade 4. (a) 2007, and (b) 1995

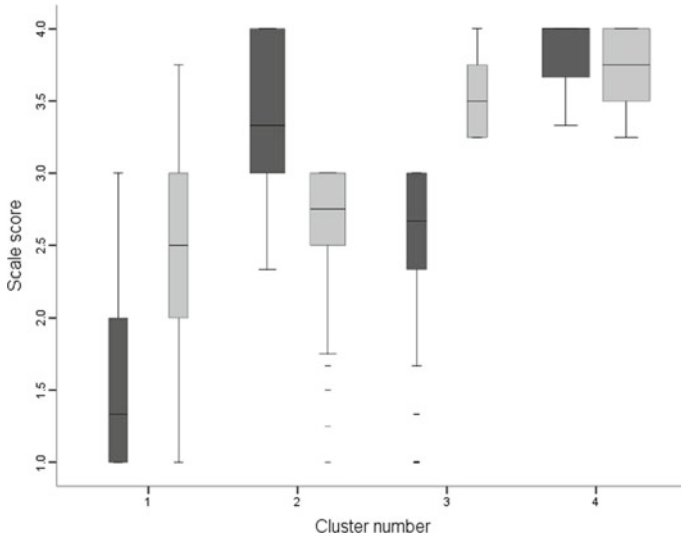


Fig. B.9 Distributions of motivation variables by cluster for Norway at grade 4 in 2007. No data were available for 1995

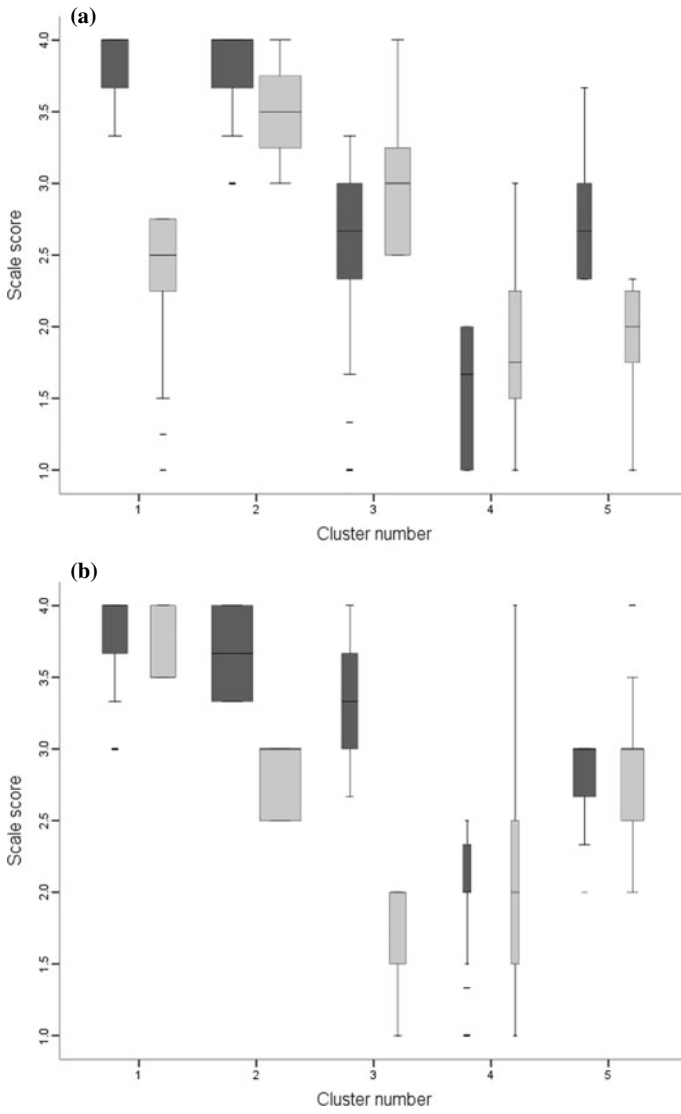


Fig. B.10 Distributions of motivation variables by cluster for Singapore at grade 4. (a) 2007, and (b) 1995

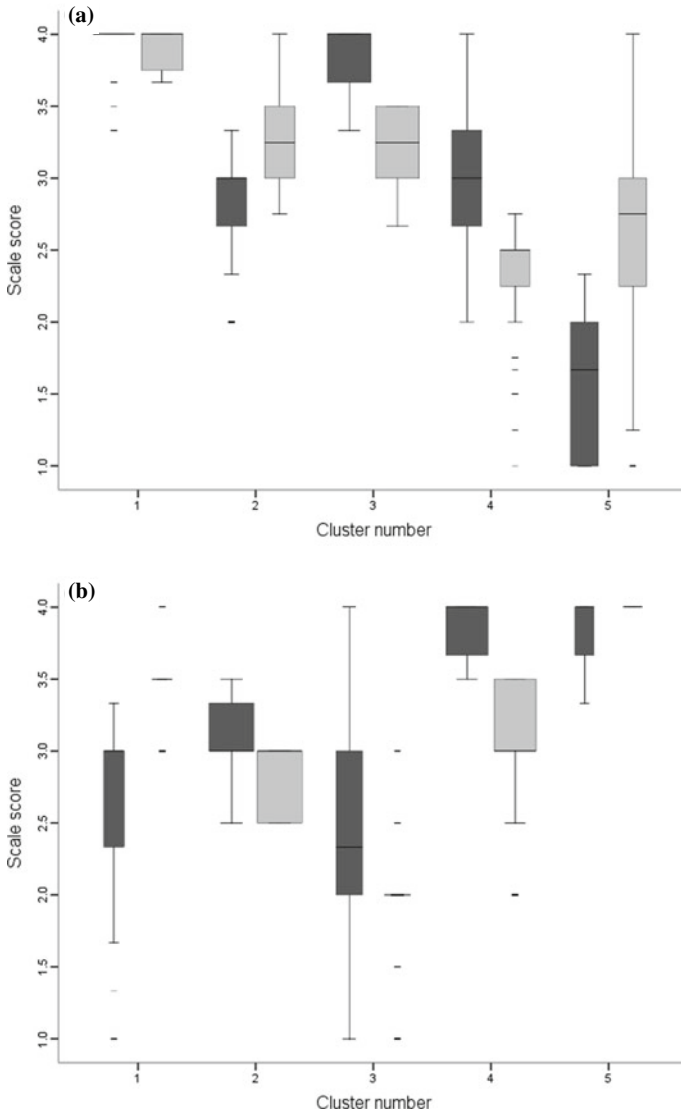


Fig. B.11 Distributions of motivation variables by cluster for Slovenia at grade 4. (a) 2007, and (b) 1995

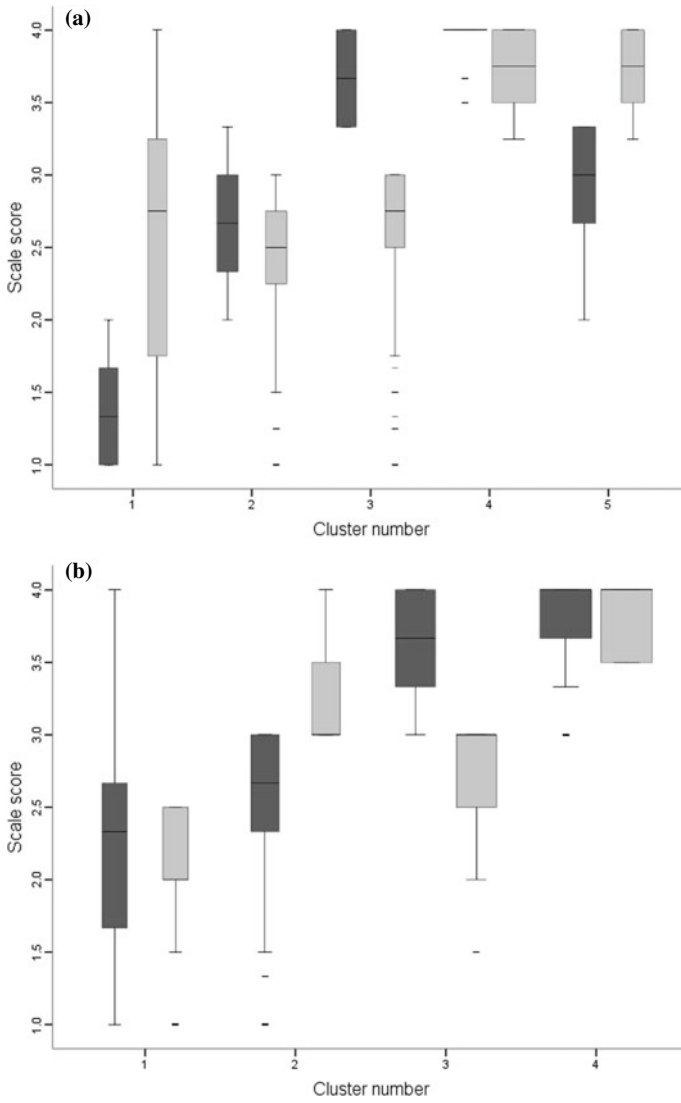


Fig. B.12 Distributions of motivation variables by cluster for the United States at grade 4. (a) 2007, and (b) 1995

B.2 Grade Eight

See Figs. B.13, B.14, B.15, B.16, B.17, B.18, B.19, B.20, B.21, B.22, B.23, and B.24.

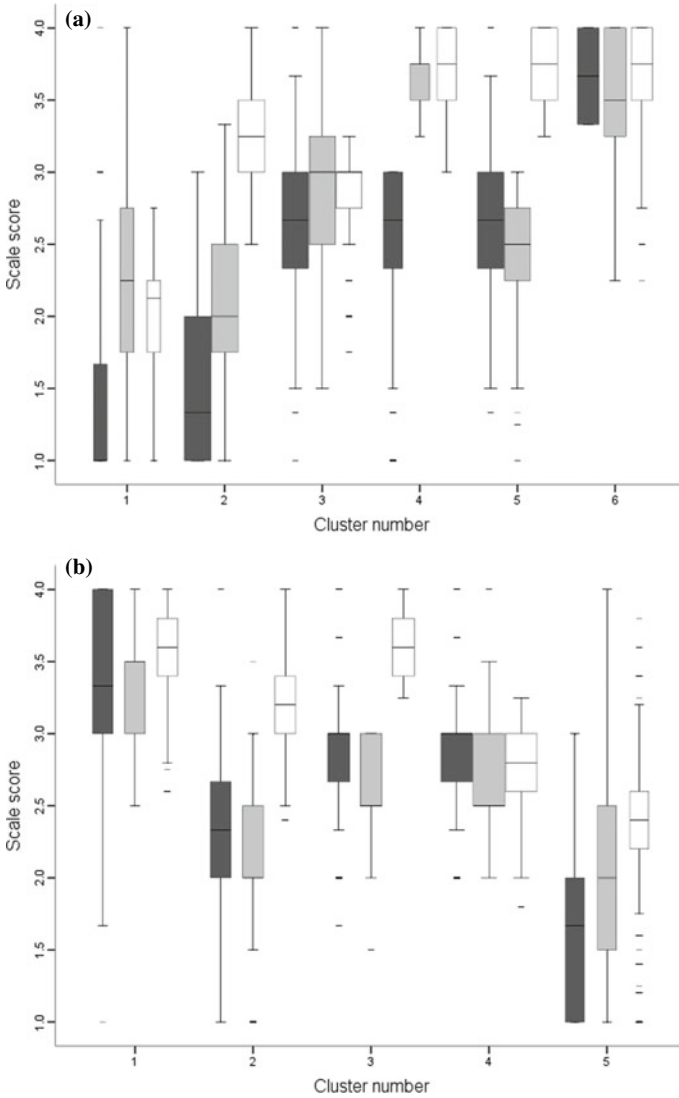


Fig. B.13 Distributions of motivation variables by cluster for Australia at grade 8. (a) 2007, and (b) 1995

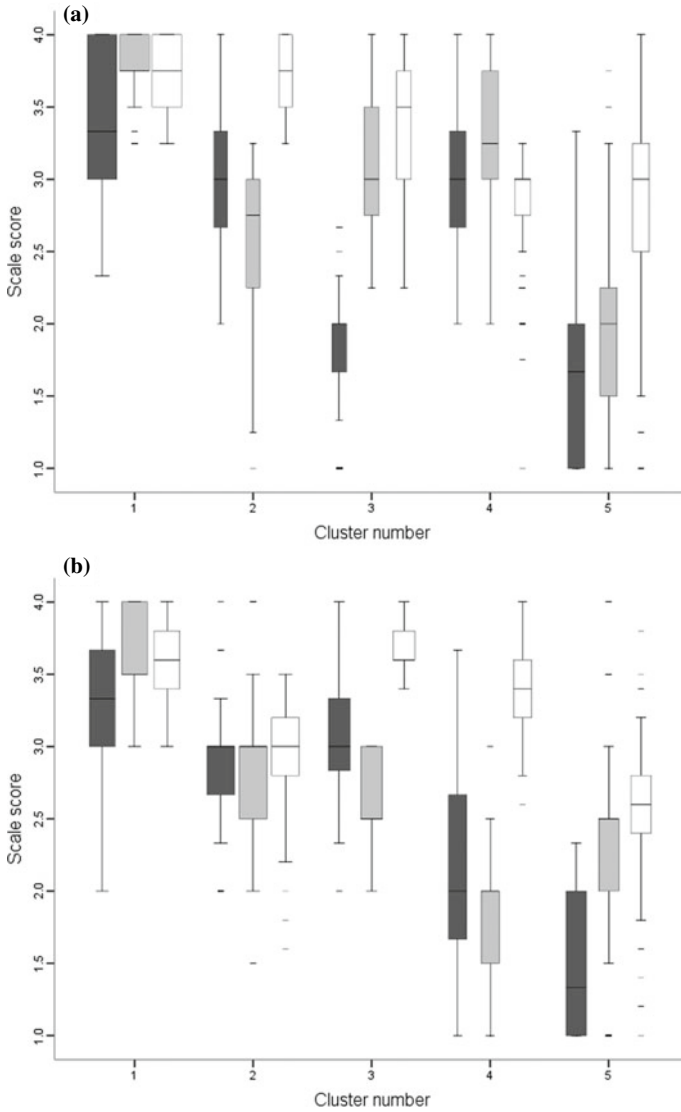


Fig. B.14 Distributions of motivation variables by cluster for Canada-Ontario at grade 8. (a) 2007, and (b) 1995

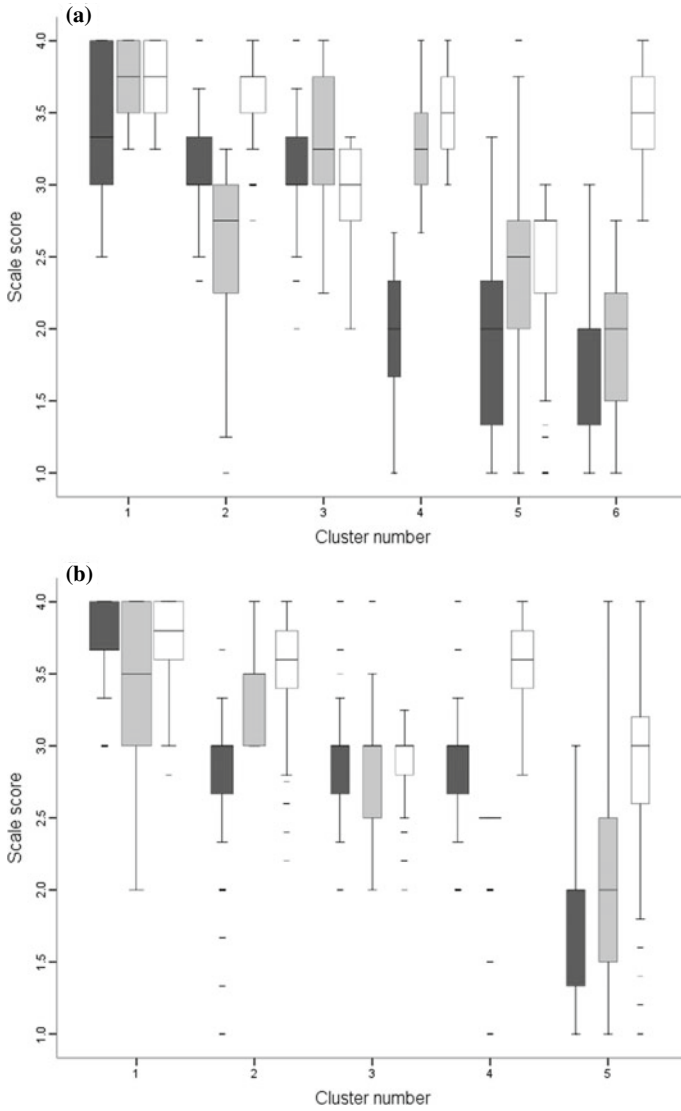


Fig. B.15 Distributions of motivation variables by cluster for Canada-Quebec at grade 8. (a) 2007, and (b) 1995

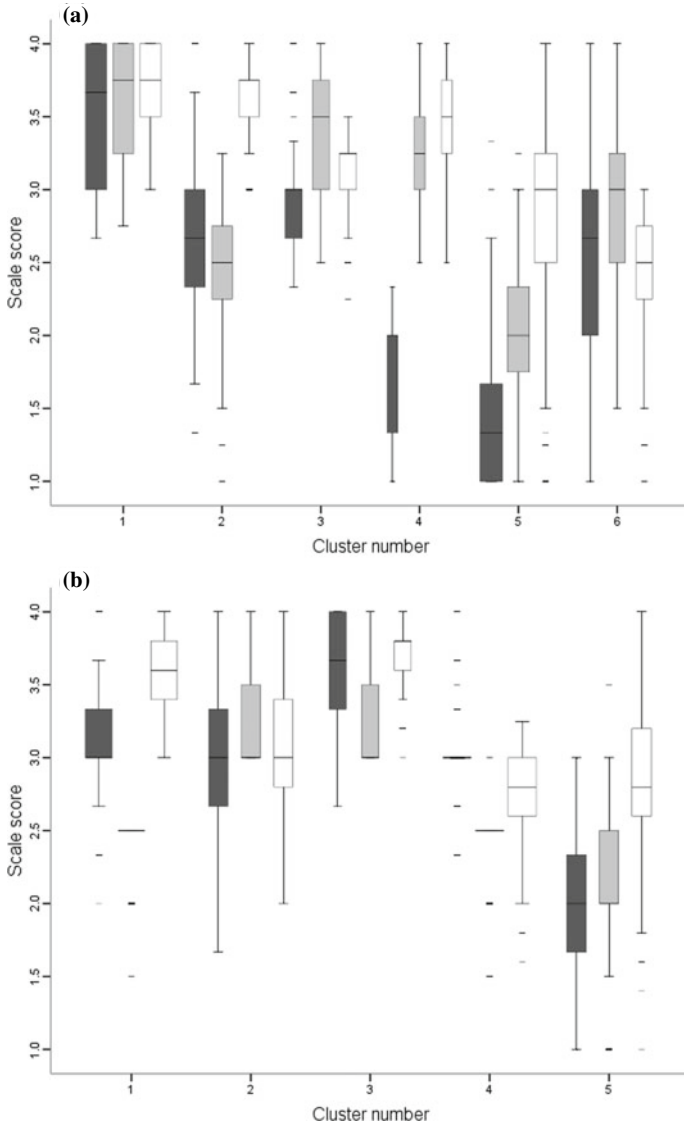


Fig. B.16 Distributions of motivation variables by cluster for England at grade 8. (a) 2007, and (b) 1995

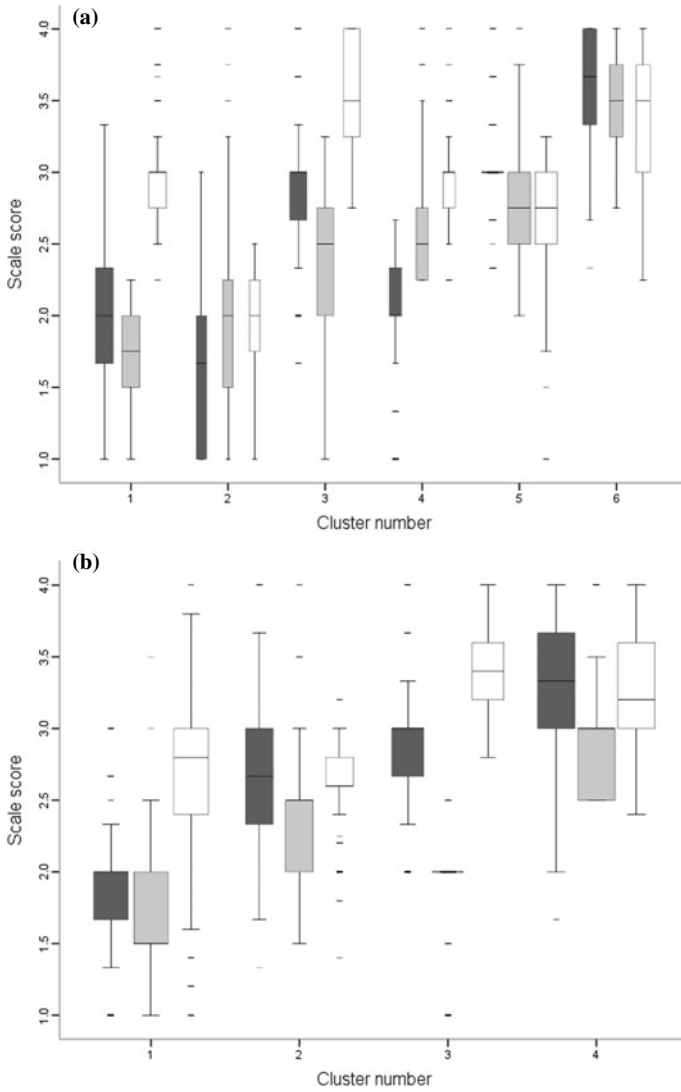


Fig. B.17 Distributions of motivation variables by cluster for Hong Kong at grade 8. (a) 2007, and (b) 1995

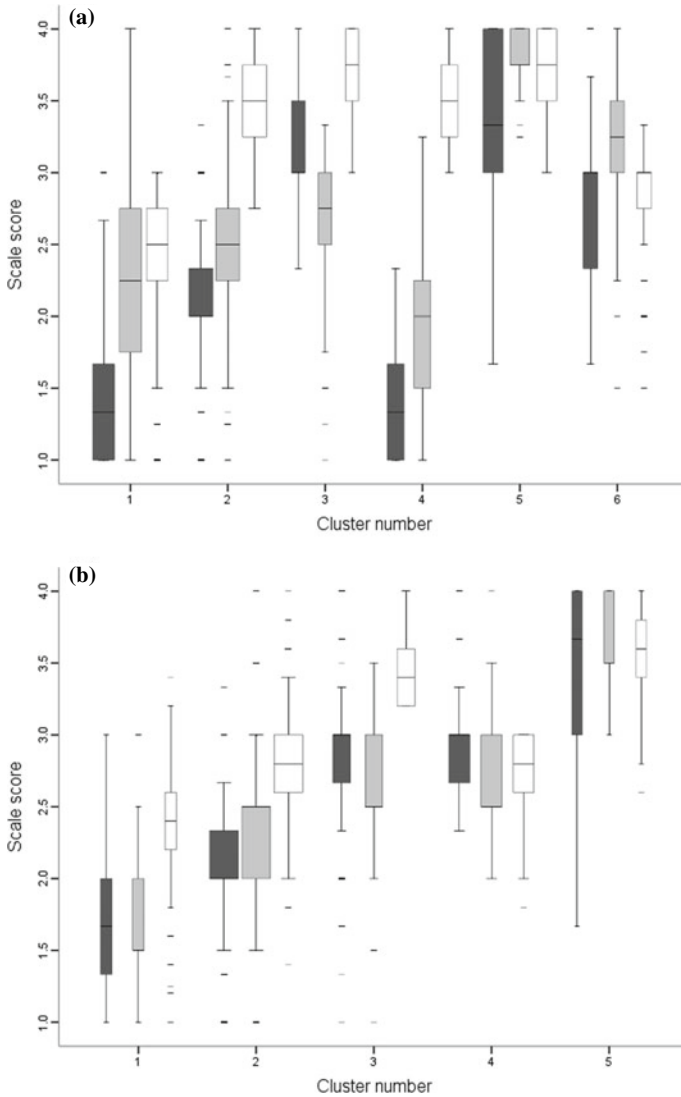


Fig. B.18 Distributions of motivation variables by cluster for Hungary at grade 8. (a) 2007, and (b) 1995

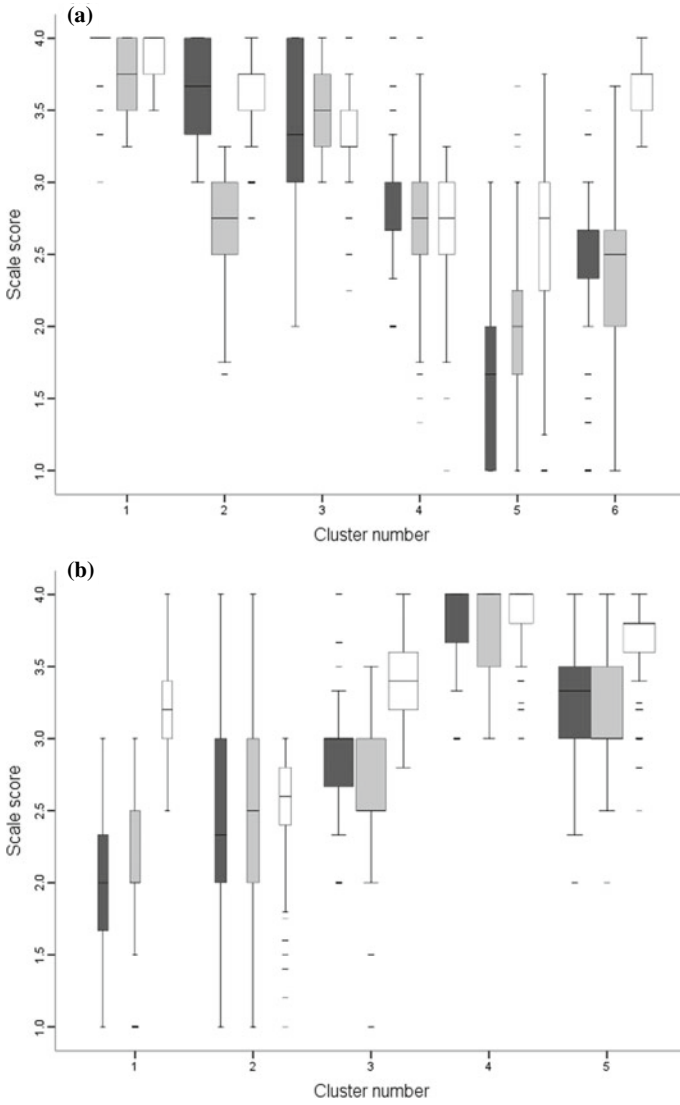


Fig. B.19 Distributions of motivation variables by cluster for Iran at grade 8. (a) 2007, and (b) 1995

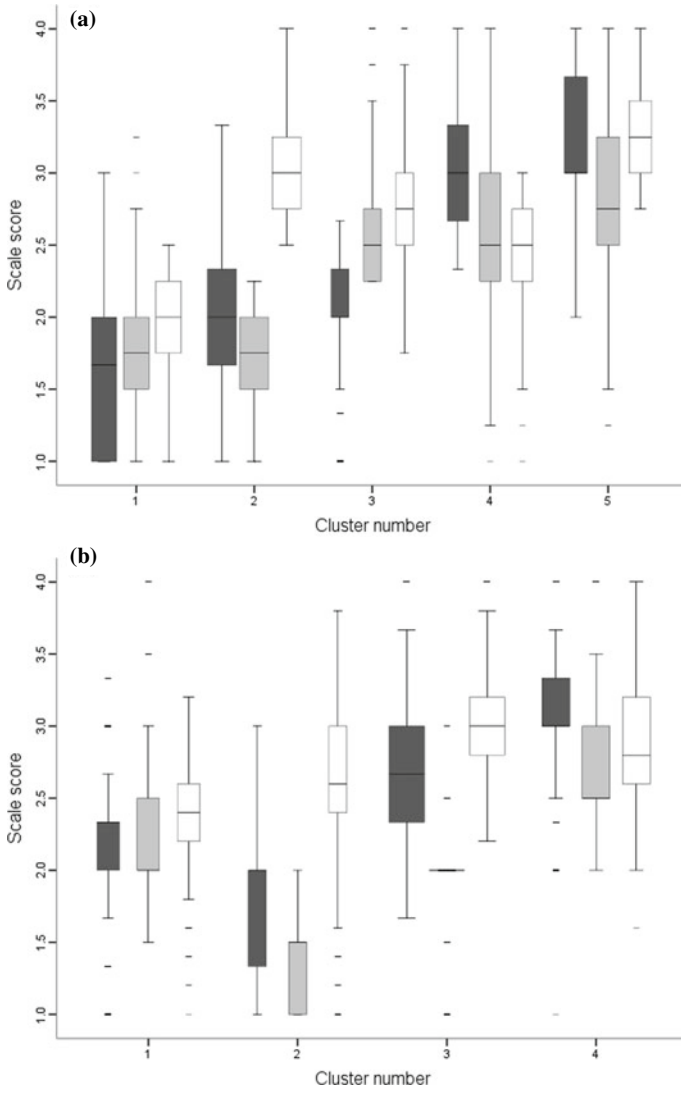


Fig. B.20 Distributions of motivation variables by cluster for Japan at grade 8. (a) 2007, and (b) 1995

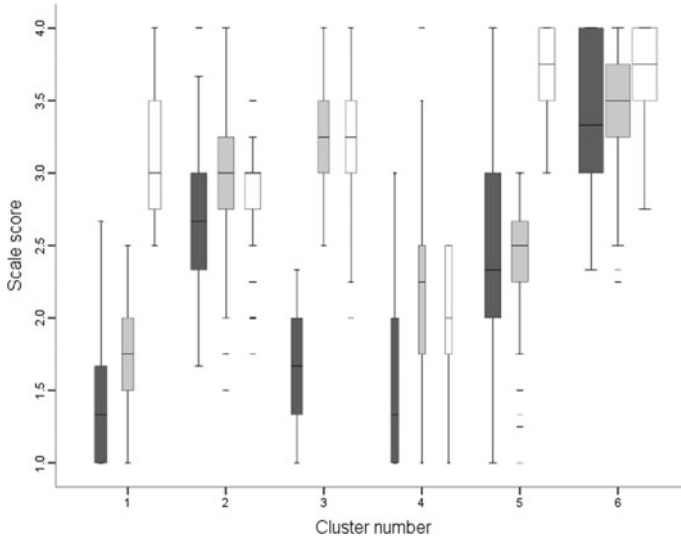


Fig. B.21 Distributions of motivation variables by cluster for Norway at grade 8 for 2007. No data were available for 1995

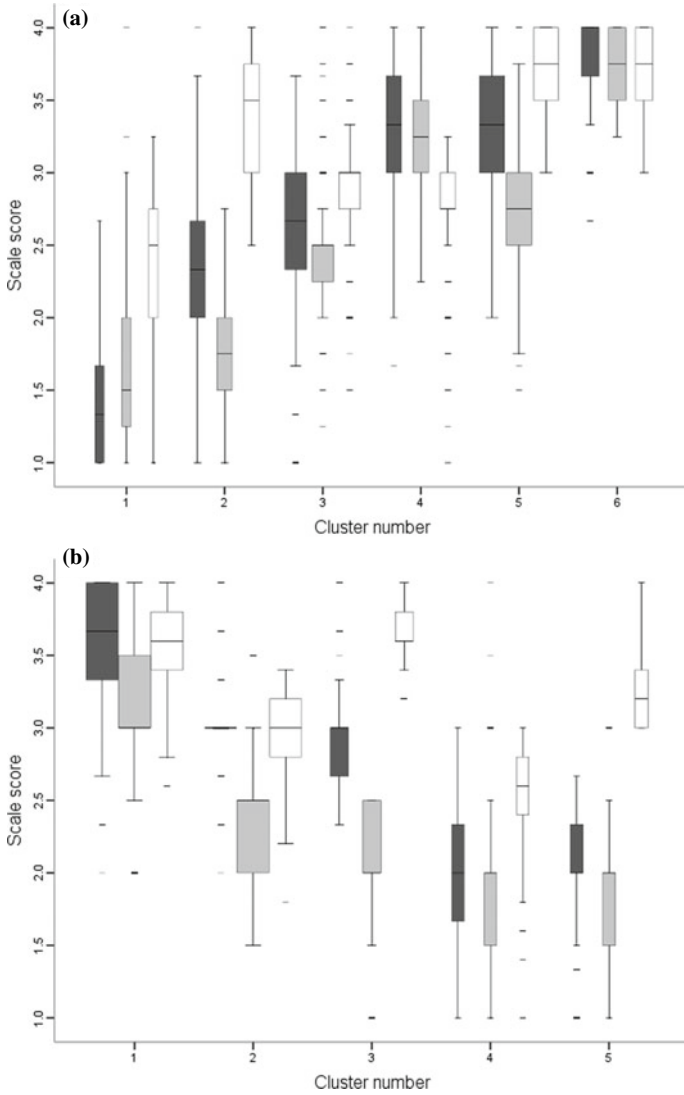


Fig. B.22 Distributions of motivation variables by cluster for Singapore at grade 8. (a) 2007, and (b) 1995

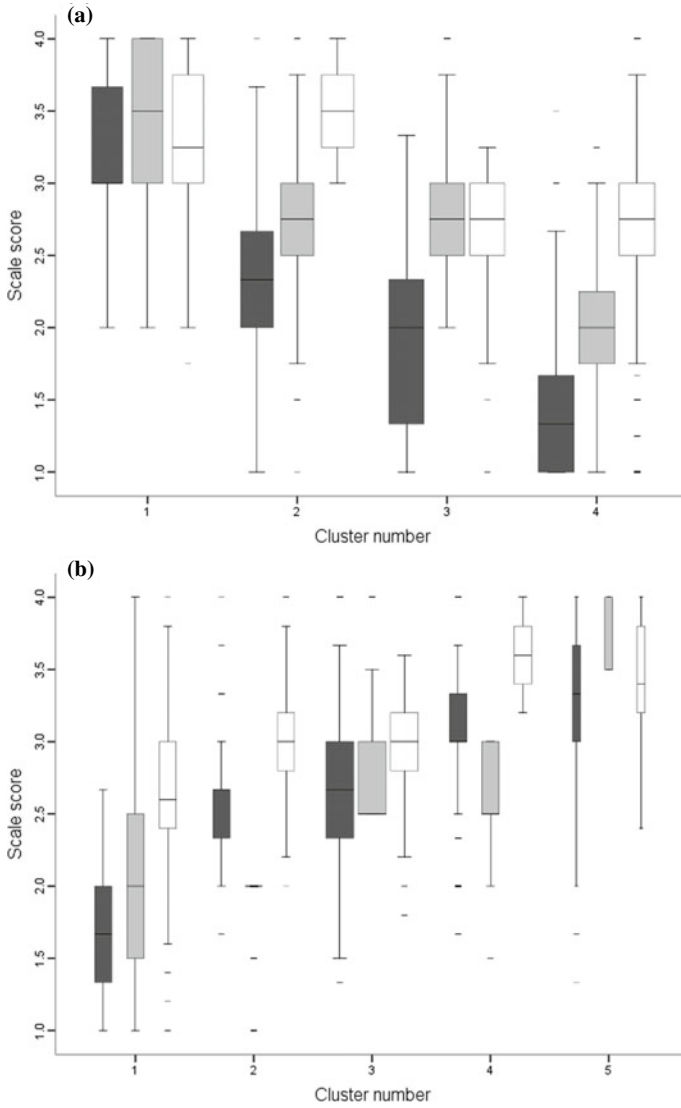


Fig. B.23 Distributions of motivation variables by cluster for Slovenia at grade 8. (a) 2007, and (b) 1995

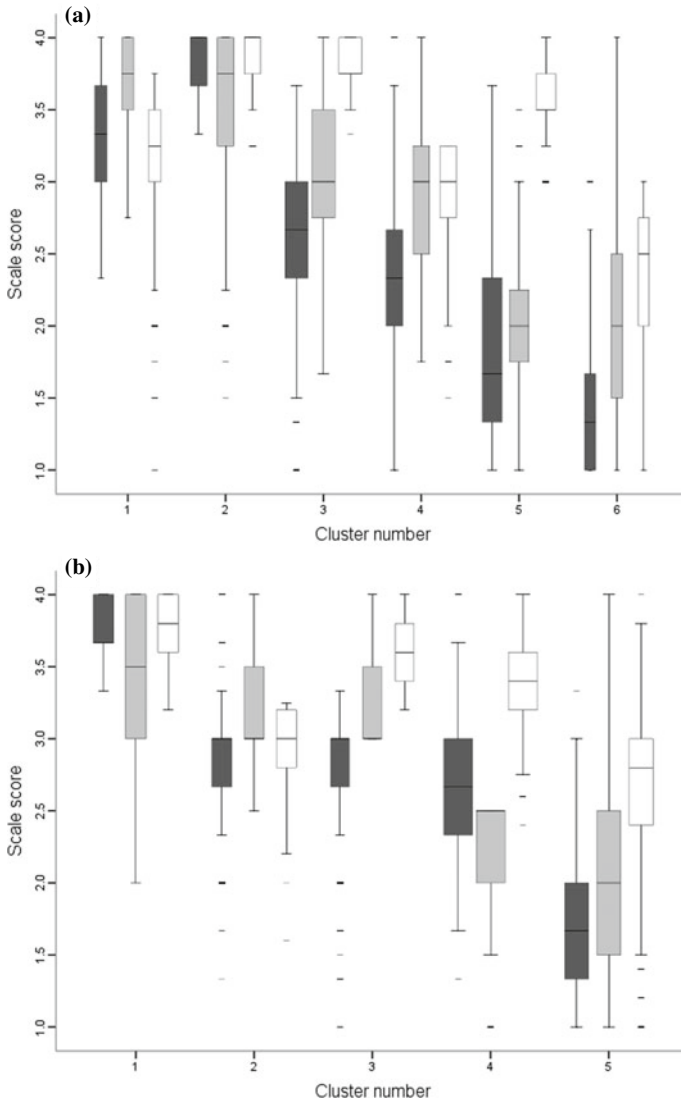


Fig. B.24 Distributions of motivation variables by cluster for the United States at grade 8. (a) 2007, and (b) 1995

Appendix C

TIMSS 1995 and 2007 Descriptive Statistics by Cluster for Each Jurisdiction

In Chap. 4, we presented results by jurisdiction for the TIMSS 2015 grade four and grade eight samples in alphabetical order by country. Scale scores were derived after conducting advanced latent variable methodology for the motivation variables, and for the home resources measure available in the TIMSS 2015 datasets (unavailable in previous administrations). In TIMSS 2015, motivation scales contained more items (see Chap. 3). In the TIMSS 2007 and 1995 cycles, simpler procedures were used for scoring: averages of items were estimated to generate the motivation variables, and single items were used as proxies of socioeconomic background. Distributions of motivation variables by cluster and jurisdiction for the TIMSS 2007 and 1995 administrations are provided in Appendix B. We present summaries of the cluster analysis results for all administrations and grades, along with trend comparisons, in Chap. 5.

Here we provide descriptive statistics with statistical tests by cluster and jurisdiction for the TIMSS 1995 and 2007 administrations. In each case, cluster quality was assessed using the silhouette procedure available in the IBM SPSS Statistics software. Descriptive statistics by cluster are presented for cluster size, mean achievement, gender composition, time spent on homework, and parental education level (only for grade eight). In all tables presented in this section, the procedure “Percentages and Means” in the IEA IDB Analyzer was used to obtain weighted means by cluster for PVs (student weight was used). Note that different superscripts (a, b, c, etc.) on the tables indicate significantly different mean plausible values, or home resources for learning based on t-statistics in pairwise comparisons. Due to multiple comparisons conducted in each sample, a difference was considered significant if $|t| > 3.29$.

C.1 Grade Four, TIMSS 1995

See Tables C.1, C.2, C.3, C.4, C.5, C.6, C.7, C.8, C.9, C.10, and C.11. Note that there are no grade four data for Norway for the 1995 cycle of TIMSS.

Table C.1 Descriptive statistics by cluster for Australia, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	16.9	10.5	25.7	20.7	26.3
Mean plausible value	496.4 ^a	517.3 ^{a,b}	521.0 ^b	543.4 ^c	549.0 ^c
Female students in cluster (%) [*]	46.5	41.7	61.1	53.2	45.0
Students spending ≥ 1 h studying mathematics daily (%)	22.7	19.3	24.4	30.1	26.1
Cases excluded, as a percentage (and number)	7.0% (426)				

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 110.862, p < 0.001$).

Table C.2 Descriptive statistics by cluster for Canada-Ontario, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	12.4	29.0	26.6	32.1
Mean plausible value	541.8 ^a	567.4 ^a	570.3 ^{a,b}	600.2 ^b
Female students in cluster (%)	51.1	43.7	52.1	40.4
Students spending ≥ 1 h studying mathematics daily (%)	20.0	26.5	31.4	26.2
Cases excluded, as a percentage (and number)	1.7% (12)			

Notes ^{*}Chi-square test of independence of gender \times cluster was not significant ($\chi^2(3) = 7.139, p = 0.068$)

Table C.3 Descriptive statistics by cluster for Canada-Quebec, TIMSS 1995 grade 4. Cluster quality is assessed as fair

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	8.9	26.1	27.9	37.1
Mean plausible value	508.2 ^a	515.8 ^a	508.4 ^a	532.1 ^b
Female students in cluster (%) [*]	43.5	51.9	55.0	48.8
Students spending ≥ 1 h studying mathematics daily (%)	21.2	33.1	26.6	27.5
Cases excluded, as a percentage (and number)	3.2% (142)			

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 19.246, p < 0.001$).

Table C.4 Descriptive statistics by cluster for England, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	23.0	51.0	11.6	14.4
Mean plausible value	512.2 ^{a,b}	523.9 ^a	495.7 ^b	504.5 ^{a,b}
Female students in cluster (%) [*]	54.1	51.6	43.6	46.4
Students spending ≥ 1 h studying mathematics daily (%)	–	–	–	–
Cases excluded, as a percentage (and number)	5.6% (175)			

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 13.599, p = 0.004$).

Table C.5 Descriptive statistics by cluster for Hong Kong, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	15.0	32.6	22.5	13.4	16.6
Mean plausible value	561.1 ^a	586.8 ^b	601.4 ^c	573.6 ^{a,b}	607.6 ^c
Female students in cluster (%) [*]	55.1	48.6	42.8	52.6	31.1
Students spending ≥ 1 h studying mathematics daily (%)	47.0	49.7	54.3	53.0	47.6
Cases excluded, as a percentage (and number)	1.0% (44)				

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 101.790, p < 0.001$).

Table C.6 Descriptive statistics by cluster for Hungary, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	14.0	27.9	27.6	30.5
Mean plausible value	533.7 ^a	535.9 ^a	538.7 ^a	581.7 ^b
Female students in cluster (%) [*]	46.3	55.5	48.5	48.4
Students spending ≥ 1 h studying mathematics daily (%)	31.9	38.0	35.8	37.8
Cases excluded, as a percentage (and number)	3.7% (110)			

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 13.678, p = 0.003$).

Table C.7 Descriptive statistics by cluster for Iran, TIMSS 1995 grade 4. Cluster quality is assessed as fair/good

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	30.9	26.2	27.0	15.9
Mean plausible value	448.2 ^a	449.9 ^a	417.2 ^b	419.5 ^b
Female students in cluster (%)*	53.5	50.9	46.7	46.5
Students spending ≥ 1 h studying mathematics daily (%)	83.2	80.4	75.3	70.4
Cases excluded, as a percentage (and number)	18.2% (615)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 9.666, p = 0.022$).

Table C.8 Descriptive statistics by cluster for Japan, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	17.4	21.6	32.4	14.6	14.0
Mean plausible value	641.5 ^a	611.7 ^b	600.2 ^b	555.4 ^c	558.0 ^c
Female students in cluster (%)*	38.5	46.2	52.3	60.2	54.3
Students spending ≥ 1 h studying mathematics daily (%)	48.3	31.4	27.7	27.6	19.3
Cases excluded, as a percentage (and number)	0.8% (34)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 77.929, p < 0.001$).

Table C.9 Descriptive statistics by cluster for Singapore, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	22.1	36.7	14.0	7.2	20.0
Mean plausible value	666.0 ^a	642.6 ^b	580.1 ^c	569.6 ^c	608.9 ^d
Female students in cluster (%)*	41.3	50.1	53.7	38.7	48.6
Students spending ≥ 1 h studying mathematics daily (%)	–	–	–	–	–
Cases excluded, as a percentage (and number)	1.5% (110)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 62.285, p < 0.001$).

Table C.10 Descriptive statistics by cluster for Slovenia, TIMSS 1995 grade 4. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	13.4	29.4	17.4	27.6	12.3
Mean plausible value	553.9 ^{a,b}	546.6 ^{b,c}	533.6 ^b	561.4 ^{a,c}	581.4 ^a
Female students in cluster (%) [*]	44.0	54.0	54.1	50.5	45.0
Students spending ≥ 1 h studying mathematics daily (%)	35.6	39.7	37.1	46.8	38.9
Cases excluded, as a percentage (and number)	2.0% (52)				

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 15.258, p = 0.004$)

Table C.11 Descriptive statistics by cluster for the United States, TIMSS 1995 grade 4. Cluster quality is assessed as good

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	17.4	19.5	27.6	35.6
Mean plausible value	519.4 ^a	544.5 ^b	549.7 ^b	556.9 ^b
Female students in cluster (%) [*]	53.1	49.6	55.2	49.1
Students spending ≥ 1 h studying mathematics daily (%)	35.3	30.0	37.3	36.2
Cases excluded, as a percentage (and number)	2.1% (150)			

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 20.070, p < 0.001$)

C.2 Grade Eight, TIMSS 1995

See Tables C.12, C.13, C.14, C.15, C.16, C.17, C.18, C.19, C.20, C.21, and C.22. Note that there are no grade eight data for Norway from the 1995 cycle of TIMSS.

Table C.12 Descriptive statistics by cluster for Australia, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	17.8	18.2	19.4	28.0	16.6
Mean plausible value	555.8 ^a	489.0 ^b	514.3 ^c	519.8 ^c	476.6 ^b
Female students in cluster (%) [*]	44.9	55.3	47.1	57.4	50.7
Parental education level post-secondary and above (%)	71.8	64.5	69.5	65.3	54.9
Students spending >1 h on homework daily (%)	25.8	20.8	27.0	18.5	9.7
Cases excluded, as a percentage (and number)	5.7% (425)				

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 66.821, p < 0.001$)

Table C.13 Descriptive statistics by cluster for Canada-Ontario, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	22.8	24.3	19.2	15.9	17.8
Mean plausible value	588.4 ^a	561.0 ^b	555.3 ^{a,b}	523.1 ^c	543.4 ^{b,c}
Female students in cluster (%) [*]	42.4	54.7	47.7	56.0	48.9
Parental education level post-secondary and above (%)	78.8	69.8	83.8	76.3	73.1
Students spending >1 h on homework daily (%)	22.9	15.9	29.5	31.3	7.6
Cases excluded, as a percentage (and number)	1.4% (15)				

Notes ^{*}Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 10.592, p = 0.032$)

Table C.14 Descriptive statistics by cluster for Canada-Quebec, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	27.0	20.0	16.6	19.9	16.5
Mean plausible value	530.8 ^a	529.9 ^a	497.8 ^b	472.8 ^c	464.1 ^c
Female students in cluster (%) [*]	44.7	46.1	53.9	54.6	54.9
Parental education level post-secondary and above (%)	77.0	80.7	72.3	69.8	70.5
Students spending >1 h on homework daily (%)	33.1	20.4	21.5	41.1	17.6
Cases excluded, as a percentage (and number)	3.3% (140)				

Notes ^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 33.710$, $p < 0.001$)

Table C.15 Descriptive statistics by cluster for England, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	24.1	17.3	15.5	25.6	17.5
Mean plausible value	497.6 ^{a,b}	504.9 ^{a,b}	516.8 ^a	502.9 ^{a,b}	483.9 ^b
Female students in cluster (%) [*]	42.5	46.9	38.8	55.5	55.3
Parental education level post-secondary and above (%)	–	–	–	–	–
Students spending >1 h on homework daily (%)	–	–	–	–	–
Cases excluded, as a percentage (and number)	5.3% (95)				

Notes ^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 29.592$, $p < 0.001$)

Table C.16 Descriptive statistics by cluster for Hong Kong, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	26.4	21.0	24.2	28.4
Mean plausible value	534.7 ^a	570.0 ^b	574.3 ^b	601.9 ^c
Female students in cluster (%) [*]	57.9	46.6	46.6	31.3
Parental education level post-secondary and above (%)	36.0	37.0	40.5	39.6
Students spending >1 h on homework daily (%)	24.4	30.8	43.6	37.2
Cases excluded, as a percentage (and number)	2.7% (90)			

Notes ^{*}Chi-square test of independence of gender × cluster was significant ($\chi^2(3) = 128.751$, $p < 0.001$)

Table C.17 Descriptive statistics by cluster for Hungary, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	12.7	32.5	19.8	23.6	11.4
Mean plausible value	489.1 ^a	510.2 ^b	544.8 ^c	530.2 ^c	591.8 ^d
Female students in cluster (%)*	47.4	53.6	52.2	53.1	47.8
Parental education level post-secondary and above (%)	88.4	89.4	91.9	88.8	94.4
Students spending >1 h on homework daily (%)	20.1	26.4	35.3	28.3	28.7
Cases excluded, as a percentage (and number)	6.3% (182)				

Notes Chi-square test of independence of gender \times cluster was not significant ($\chi^2(4) = 6.360, p = 0.174$)

Table C.18 Descriptive statistics by cluster for Iran, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	9.5	10.9	27.5	22.2	29.9
Mean plausible value	386.9 ^a	402.1 ^a	412.5 ^{a,b}	451.3 ^c	428.4 ^b
Female students in cluster (%)*	47.9	46.2	45.1	44.7	41.6
Parental education level post-secondary and above (%)	19.8	25.6	21.7	32.4	26.6
Students spending >1 h on homework daily (%)	64.0	68.0	77.7	87.7	81.5
Cases excluded, as a percentage (and number)	10.7% (394)				

Notes Chi-square test of independence of gender \times cluster was not significant ($\chi^2(4) = 5.432, p = 0.246$)

Table C.19 Descriptive statistics by cluster for Japan, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster			
	1	2	3	4
Size (% of total number of students)	21.5	17.4	34.3	26.9
Mean plausible value	580.4 ^a	534.9 ^b	572.0 ^a	623.3 ^c
Female students in cluster (%)*	51.5	59.9	47.8	39.9
Parental education level post-secondary and above (%)	–	–	–	–
Students spending >1 h on homework daily (%)	21.8	21.0	35.7	35.1
Cases excluded, as a percentage (and number)	0.7% (35)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 91.295, p < 0.001$)

Table C.20 Descriptive statistics by cluster for Singapore, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	30.0	29.8	16.9	11.4	11.9
Mean plausible value	627.0 ^a	613.5 ^b	606.7 ^b	575.3 ^c	582.6 ^c
Female students in cluster (%)*	43.5	52.5	53.0	52.5	51.7
Parental education level post-secondary and above (%)	77.9	74.0	77.8	71.4	74.5
Students spending >1 h on homework daily (%)	79.7	76.0	84.1	61.2	70.7
Cases excluded, as a percentage (and number)	1.0% (47)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 31.237, p < 0.001$)

Table C.21 Descriptive statistics by cluster for Slovenia, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	19.3	19.6	32.1	20.0	8.9
Mean plausible value	510.6 ^a	496.1 ^b	538.5 ^c	543.1 ^c	601.4 ^d
Female students in cluster (%)*	50.5	50.2	53.7	52.2	43.3
Parental education level post-secondary and above (%)	76.6	70.1	81.5	79.7	87.4
Students spending >1 h on homework daily (%)	26.8	46.9	32.2	43.6	19.3
Cases excluded, as a percentage (and number)	1.1% (30)				

Notes Chi-square test of independence of gender \times cluster was not significant ($\chi^2(4) = 8.622, p = 0.071$)

Table C.22 Descriptive statistics by cluster for the United States, TIMSS 1995 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster				
	1	2	3	4	5
Size (% of total number of students)	18.3	17.9	16.6	25.9	21.4
Mean plausible value	519.8 ^a	498.5 ^b	507.9 ^{a,b}	483.1 ^c	470.8 ^c
Female students in cluster (%)*	49.9	51.3	43.6	53.7	51.0
Parental education level post-secondary and above (%)	87.2	85.4	87.0	85.8	82.3
Students spending >1 h on homework daily (%)	39.6	23.0	24.8	33.0	15.6
Cases excluded, as a percentage (and number)	3.2% (224)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 29.391, p < 0.001$)

C.3 Grade Four, TIMSS 2007

See Tables C.23, C.24, C.25, C.26, C.27, C.28, C.29, C.30, C.31, C.32, C.33 and C.34.

Table C.23 Descriptive statistics by cluster for Australia, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster			
		1	2	3	4
Size (% of total number of students)		13.7	28.7	20.7	36.9
Mean plausible value		468.9 ^a	482.8 ^a	542.6 ^b	547.7 ^b
Female students in cluster (%)*		51.1	59.9	46.9	44.1
Time spent on mathematics homework (TIMSS index)	High	8.3	8.2	6.3	5.5
	Medium	35.4	42.8	42.7	44.2
	Low	56.3	49.0	50.9	50.3
Cases excluded, as a percentage (and number)		2.0% (82)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 69.098, p < 0.001$)

Table C.24 Descriptive statistics by cluster for Canada-Ontario, TIMSS 2007 grade 4. Cluster quality is assessed as fair

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		16.0	15.4	18.0	27.9	22.7
Mean plausible value		481.8 ^c	538.3 ^a	476.8 ^c	509.6 ^b	551.3 ^a
Female students in cluster (%)*		45.4	46.2	61.9	51.7	41.5
Time spent on mathematics homework (TIMSS index)	High	14.1	10.9	11.2	11.0	9.8
	Medium	53.7	51.6	54.6	54.2	51.0
	Low	32.2	37.5	34.3	34.8	39.2
Cases excluded, as a percentage (and number)		2.1% (74)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 65.178, p < 0.001$)

Table C.25 Descriptive statistics by cluster for Canada-Quebec, TIMSS 2007 grade 4. Cluster quality is assessed as fair/good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		24.1	25.4	21.0	14.9	14.5
Mean plausible value		560.7 ^a	511.8 ^b	484.9 ^c	551.1 ^a	484.1 ^c
Female students in cluster (%)*		42.4	50.3	58.8	49.5	58.0
Time spent on mathematics homework (TIMSS index)	High	5.5	9.3	9.6	6.1	10.4
	Medium	47.2	48.1	48.3	41.7	46.1
	Low	47.2	42.6	42.1	52.2	43.4
Cases excluded, as a percentage (and number)		2.2% (84)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 58.231, p < 0.001$)

Table C.26 Descriptive statistics by cluster for England, TIMSS 2007 grade 4. Cluster quality is assessed as fair/good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		14.0	21.6	25.0	20.1	19.3
Mean plausible value		510.6 ^c	502.5 ^c	533.8 ^b	578.0 ^a	582.1 ^a
Female students in cluster (%)*		52.5	60.0	49.3	48.9	39.6
Time spent on mathematics homework (TIMSS index)	High	1.5	2.9	3.0	2.2	3.1
	Medium	27.7	31.0	31.3	28.1	31.1
	Low	70.9	66.0	65.7	69.7	65.8
Cases excluded, as a percentage (and number)		1.4 (60)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 74.514, p < 0.001$)

Table C.27 Descriptive statistics by cluster for Hong Kong, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		19.9	18.7	26.8	13.7	20.9
Mean plausible value		577.3 ^d	628.2 ^b	587.3 ^d	604.4 ^c	642.9 ^a
Female students in cluster (%)*		56.8	44.5	57.9	47.6	33.1
Time spent on mathematics homework (TIMSS index)	High	27.2	14.2	20.9	16.4	10.2
	Medium	67.9	82.7	73.3	79.6	87.9
	Low	4.9	3.1	5.7	4.0	1.9
Cases excluded, as a percentage (and number)		0.8% (29)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 135.224, p < 0.001$)

Table C.28 Descriptive statistics by cluster for Hungary, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		27.9	14.9	15.4	20.7	21.1
Mean plausible value		560.2 ^a	561.5 ^a	488.7 ^b	486.6 ^b	464.6 ^c
Female students in cluster (%)*		45.5	48.1	44.9	55.5	54.8
Time spent on mathematics homework (TIMSS index)	High	16.8	19.7	26.7	21.7	24.5
	Medium	80.0	76.1	69.5	72.5	70.7
	Low	3.2	4.2	3.8	5.7	4.8
Cases excluded, as a percentage (and number)		1.8% (74)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 33.564, p < 0.001$)

Table C.29 Descriptive statistics by cluster for Iran, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster			
		1	2	3	4
Size (% of total number of students)		29.5	28.3	27.8	14.4
Mean plausible value		458.5 ^a	429.6 ^b	388.2 ^c	359.1 ^d
Female students in cluster (%)*		49.9	43.1	50.7	47.0
Time spent on mathematics homework (TIMSS index)	High	38.4	39.0	32.9	30.1
	Medium	49.0	50.9	51.3	50.3
	Low	12.6	10.1	15.8	19.7
Cases excluded, as a percentage (and number)		25.6% (983)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 11.394, p = 0.010$)

Table C.30 Descriptive statistics by cluster for Japan, TIMSS 2007 grade 4. Cluster quality is assessed as fair/good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		34.9	18.5	16.5	21.2	9.0
Mean plausible value		600.7 ^a	591.7 ^a	547.8 ^b	531.6 ^c	527.0 ^c
Female students in cluster (%)*		40.2	43.3	58.4	59.6	56.2
Time spent on mathematics homework (TIMSS index)	High	7.1	7.5	12.5	15.1	16.5
	Medium	67.0	65.5	61.4	59.3	59.4
	Low	25.9	27.0	26.1	25.6	24.2
Cases excluded, as a percentage (and number)		1.7% (76)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 133.186, p < 0.001$)

Table C.31 Descriptive statistics by cluster for Norway, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster			
		1	2	3	4
Size (% of total number of students)		15.5	29.0	16.5	39.1
Mean plausible value		454.0 ^b	445.1 ^b	497.5 ^a	495.6 ^a
Female students in cluster (%)*		45.6	54.7	47.1	48.0
Time spent on mathematics homework (TIMSS index)	High	16.5	12.5	5.6	8.4
	Medium	50.4	51.4	54.1	51.4
	Low	33.1	36.0	40.3	40.3
Cases excluded, as a percentage (and number)		2.7% (112)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 19.546, p < 0.001$)

Table C.32 Descriptive statistics by cluster for Singapore, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		22.2	34.6	20.6	10.6	12.1
Mean plausible value		574.7 ^b	640.4 ^a	612.8 ^c	557.9 ^{b,d}	544.6 ^d
Female students in cluster (%)*		56.4	41.3	50.7	47.6	59.4
Time spent on mathematics homework (TIMSS index)	High	32.7	33.2	35.4	33.5	35.4
	Medium	52.6	54.0	48.2	47.9	50.9
	Low	14.7	12.8	16.3	18.6	13.8
Cases excluded, as a percentage (and number)		0.6% (31)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 93.301, p < 0.001$)

Table C.33 Descriptive statistics by cluster for Slovenia, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		24.0	17.4	25.4	17.2	16.0
Mean plausible value		541.4 ^a	518.9 ^c	499.2 ^b	450.1 ^d	488.7 ^b
Female students in cluster (%)*		46.2	49.3	51.0	56.9	44.7
Time spent on mathematics homework (TIMSS index)	High	12.2	13.8	17.7	28.0	23.7
	Medium	86.6	84.6	80.5	68.9	70.6
	Low	1.2	1.6	1.8	3.1	5.7
Cases excluded, as a percentage (and number)		1.8% (78)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 27.905, p < 0.001$)

Table C.34 Descriptive statistics by cluster for the United States, TIMSS 2007 grade 4. Cluster quality is assessed as good

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		14.9	16.7	15.8	34.6	18.0
Mean plausible value		516.5 ^d	495.1 ^c	491.6 ^c	553.5 ^a	565.0 ^b
Female students in cluster (%)*		49.2	54.3	57.0	49.3	48.8
Time spent on mathematics homework (TIMSS index)	High	15.1	13.2	13.1	9.5	9.7
	Medium	62.3	63.7	66.4	68.2	67.1
	Low	22.6	23.1	20.6	22.2	23.2
Cases excluded, as a percentage (and number)		2.1% (168)				

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 30.250, p < 0.001$)

C.4 Grade Eight, TIMSS 2007

See Tables C.35, C.36, C.37, C.38, C.39, C.40, C.41, C.42, C.43, C.44, C.45, and C.46.

Table C.35 Descriptive statistics by cluster for Australia, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		10.0	19.9	19.5	14.5	19.7	16.3
Mean plausible value		466.6 ^{a,b}	457.0 ^a	502.3 ^c	558.5 ^d	475.0 ^b	532.5 ^d
Female students in cluster (%)*		52.0	54.1	49.9	37.4	42.1	36.6
Parental education level post-secondary and above (%)		35.6	32.1	39.3	52.6	36.0	48.9
Time spent on mathematics homework (TIMSS index)	High	8.7	9.9	10.7	13.2	14.6	16.6
	Medium	41.6	40.7	43.1	46.3	42.6	44.7
	Low	49.7	49.4	46.1	40.5	42.8	38.7
Cases excluded, as a percentage (and number)		3.2% (130)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 75.747, p < 0.001$)

Table C.36 Descriptive statistics by cluster for Canada-Ontario, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster				
		1	2	3	4	5
Size (% of total number of students)		32.6	15.0	16.2	17.8	18.4
Mean plausible value		561.1 ^d	488.4 ^c	522.9 ^b	522.0 ^b	462.9 ^a
Female students in cluster (%)*		43.7	56.1	51.8	45.7	61.0
Parental education level post-secondary and above (%)		69.3	52.2	56.3	51.5	44.2
Time spent on mathematics homework (TIMSS index)	High	24.1	36.8	21.5	22.0	33.6
	Medium	59.9	48.9	58.8	58.6	53.7
	Low	16.0	14.4	19.7	19.4	12.7
Cases excluded, as a percentage (and number)		3.2% (112)				

Notes *Chi-square test of independence of gender × cluster was significant ($\chi^2(4) = 58.930, p < 0.001$)

Table C.37 Descriptive statistics by cluster for Canada-Quebec, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		18.0	19.7	16.7	10.0	17.7	17.9
Mean plausible value		573.5 ^d	516.7 ^a	555.3 ^c	558.8 ^{c,d}	505.6 ^a	489.8 ^b
Female students in cluster (%)*		43.4	52.5	49.7	38.7	50.8	57.2
Parental education level post-secondary and above (%)		68.5	56.7	57.3	64.6	49.9	52.2
Time spent on mathematics homework (TIMSS index)	High	34.1	37.9	26.7	25.9	25.7	30.7
	Medium	50.5	45.8	50.4	50.7	48.8	50.0
	Low	15.5	16.3	22.9	23.4	25.5	19.3
Cases excluded, as a percentage (and number)		3.1% (121)					

Notes *Chi-square test of independence of gender × cluster was significant ($\chi^2(5) = 47.646, p < 0.001$)

Table C.38 Descriptive statistics by cluster for England, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		19.7	19.1	15.9	9.9	20.6	14.9
Mean plausible value		547.0 ^c	485.8 ^a	551.3 ^c	524.2 ^b	475.6 ^a	519.6 ^b
Female students in cluster (%)*		36.2	55.4	48.6	40.5	67.9	54.5
Parental education level post-secondary and above (%)		–	–	–	–	–	–
Time spent on mathematics homework (TIMSS index)	High	5.5	5.8	4.4	5.5	3.2	3.7
	Medium	34.4	32.0	31.4	30.1	23.5	30.7
	Low	60.1	62.2	64.2	64.4	73.3	65.6
Cases excluded, as a percentage (and number)		2.2% (87)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 187.957, p < 0.001$)

Table C.39 Descriptive statistics by cluster for Hong Kong, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		19.0	10.9	18.2	12.9	24.1	14.8
Mean plausible value		544.9 ^a	498.1 ^b	583.1 ^{c,d}	561.2 ^{a,c}	591.6 ^d	635.8 ^e
Female students in cluster (%)*		65.9	51.6	49.2	47.6	50.5	35.3
Parental education level post-secondary and above (%)		24.2	19.0	27.6	26.6	22.1	34.7
Time spent on mathematics homework (TIMSS index)	High	39.7	33.2	40.1	36.2	30.3	29.3
	Medium	44.4	40.1	45.2	45.5	52.5	51.7
	Low	16.0	26.6	14.7	18.3	17.3	18.9
Cases excluded, as a percentage (and number)		1% (33)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 111.262, p < 0.001$)

Table C.40 Descriptive statistics by cluster for Hungary, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		19.8	21.8	12.1	15.2	17.9	13.1
Mean plausible value		493.2 ^a	500.9 ^a	517.5 ^b	472.0 ^c	590.3 ^d	541.7 ^e
Female students in cluster (%)*		50.2	52.3	52.9	52.0	41.3	53.3
Parental education level post-secondary and above (%)		42.6	41.9	46.5	37.4	57.1	47.7
Time spent on mathematics homework (TIMSS index)	High	12.3	19.3	27.4	18.2	12.1	16.1
	Medium	80.9	74.5	69.1	75.0	84.0	77.5
	Low	6.7	6.3	3.4	6.7	3.9	6.4
Cases excluded, as a percentage (and number)		1.1% (45)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 28.842, p < 0.001$)

Table C.41 Descriptive statistics by cluster for Iran, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		18.1	23.5	14.9	14.4	9.7	19.4
Mean plausible value		475.5 ^d	401.2 ^b	452.7 ^c	388.8 ^{a,b}	379.2 ^{a,b}	380.1 ^a
Female students in cluster (%)*		49.9	44.0	46.7	45.3	47.7	44.0
Parental education level post-secondary and above (%)		42.2	24.2	32.5	22.2	21.6	21.4
Time spent on mathematics homework (TIMSS index)	High	29.7	25.1	26.8	21.2	16.1	17.8
	Medium	53.8	55.0	55.6	50.4	56.1	52.4
	Low	16.4	19.9	17.6	28.4	27.7	29.7
Cases excluded, as a percentage (and number)		16.0% (636)					

Notes *Chi-square test of independence of gender \times cluster was not significant ($\chi^2(5) = 6.657, p = 0.257$)

Table C.42 Descriptive statistics by cluster for Japan, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster					
	1	2	3	4	5	
Size (% of total number of students)	21.7	25.1	15.5	18.1	19.7	
Mean plausible value	529.2 ^a	538.6 ^a	595.1 ^c	592.1 ^c	618.2 ^d	
Female students in cluster (%)*	58.1	54.9	41.8	50.3	39.3	
Parental education level post-secondary and above (%)	42.3	48.7	56.8	49.1	64.8	
Time spent on mathematics homework (TIMSS index)	High	8.2	10.0	4.8	7.0	7.9
	Medium	34.0	38.8	35.5	35.8	32.3
	Low	57.8	51.1	59.7	57.2	59.8
Cases excluded, as a percentage (and number)	0.9% (37)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(4) = 90.795, p < 0.001$)

Table C.43 Descriptive statistics by cluster for Norway, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics	Cluster						
	1	2	3	4	5	6	
Size (% of total number of students)	13.7	18.4	12.8	8.2	18.6	28.3	
Mean plausible value	421.1 ^a	481.7 ^d	496.7 ^c	435.7 ^{a,b}	443.2 ^b	504.8 ^c	
Female students in cluster (%)*	55.4	49.4	44.4	50.5	52.1	47.1	
Parental education level post-secondary and above (%)	35.3	43.4	50.4	34.6	42.3	61.5	
Time spent on mathematics homework (TIMSS index)	High	25.3	24.0	21.6	21.4	29.4	27.2
	Medium	51.6	54.8	56.3	54.3	50.6	52.9
	Low	23.1	21.1	22.1	24.2	20.0	20.0
Cases excluded, as a percentage (and number)	3.2% (148)						

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 19.925, p = 0.001$)

Table C.44 Descriptive statistics by cluster for Singapore, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		8.7	14.7	20.8	14.6	24.2	17.0
Mean plausible value		525.1 ^a	553.3 ^c	573.2 ^b	638.3 ^d	591.9 ^e	645.2 ^d
Female students in cluster (%)*		46.5	53.2	47.2	46.5	52.1	46.3
Parental education level post-secondary and above (%)		29.4	30.1	34.9	41.7	38.7	47.2
Time spent on mathematics homework (TIMSS index)	High	32.1	42.7	42.3	39.4	45.1	39.0
	Medium	43.2	42.8	40.7	45.6	41.2	46.1
	Low	24.7	14.5	17.1	15.0	13.7	14.9
Cases excluded, as a percentage (and number)		0.4% (18)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 15.109, p = 0.010$)

Table C.45 Descriptive statistics by cluster for Slovenia, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster			
		1	2	3	4
Size (% of total number of students)		23.0	24.4	25.9	26.6
Mean plausible value		540.9 ^d	491.4 ^c	512.8 ^b	468.1 ^a
Female students in cluster (%)*		51.2	46.7	50.0	53.2
Parental education level post-secondary and above (%)		63.6	58.4	62.0	54.6
Time spent on mathematics homework (TIMSS index)	High	18.1	23.0	19.6	22.8
	Medium	72.7	64.3	62.8	53.3
	Low	9.2	12.6	17.6	23.9
Cases excluded, as a percentage (and number)		1.8% (73)			

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(3) = 8.855, p = 0.031$)

Table C.46 Descriptive statistics by cluster for the United States, TIMSS 2007 grade 8. Cluster quality is assessed as fair

Cluster characteristics		Cluster					
		1	2	3	4	5	6
Size (% of total number of students)		12.3	16.7	23.4	16.6	19.9	11.1
Mean plausible value		550.3 ^d	525.2 ^c	521.6 ^c	508.6 ^b	472.3 ^a	479.1 ^a
Female students in cluster (%)*		45.0	50.5	49.6	49.1	57.5	49.4
Parental education level post-secondary and above (%)		52.5	54.4	55.4	46.8	44.0	43.9
Time spent on mathematics homework (TIMSS index)	High	18.9	22.3	25.5	23.4	31.0	35.0
	Medium	67.9	65.5	62.3	63.7	58.7	54.7
	Low	13.2	12.1	12.2	12.9	10.3	10.2
Cases excluded, as a percentage (and number)		1.6% (116)					

Notes *Chi-square test of independence of gender \times cluster was significant ($\chi^2(5) = 41.047, p < 0.001$)