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 jurisdictions[1] 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.

---

[1]Norway 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 competed post-secondary education or above. For students in lower performing clusters the reported percentage of parents who had competed 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. [a]More girls were present in the high motivation and achievement clusters and more boys were present in the low motivation and achievement clusters. [b]There 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.