Exploring coherence between Swedish grades and TIMSS 2015

SUMMARY

International large-scale assessments (ILSAs) have become an important part of the Swedish evaluation system. It is therefore of crucial importance to validate national measures of Swedish students' achievement with their ILSA test scores. Here, we offer results from such a validation study based on Swedish students' test scores in IEA's Trends in International Mathematics and Science Study (TIMSS) 2015 in year 8, their final grades at year 9, and their national test grades at year 9. We find there is high consistency between what is measured in TIMSS and what is measured by indicators in the Swedish national assessment system.

IMPLICATIONS

- Students in Sweden who have higher grades tend to score higher on TIMSS in both mathematics and science, indicating that students' abilities as measured by TIMSS correspond relatively well with students' abilities as measured by their final grades.
- Correlations between students' grades and their TIMSS scores are moderately high, for both final grades and the national assessment grades, providing further evidence that the evaluation system is robust. An exact correlation should not be expected given that the curriculum and what is measured by TIMSS do not perfectly align.
- Since many reforms in the school system are based on results from ILSA, it is important to confirm that the results from the studies are consistent with what is being taught and assessed in the national system. Knowing that the consistency is moderately high legitimizes the use of results from ILSA in shaping the school system when appropriate.

INTRODUCTION

Sweden participates in several international large-scale assessments (ILSAs) which compare students' abilities in various subjects between countries and over time. The information and results gathered from them are an important part of the Swedish evaluation system. To ensure that they are relevant for such a purpose, analysis of the coherence between Swedish students' abilities as estimated in international studies and their abilities as estimated by national measures of achievement ought to be conducted. Using data collected by the International Association for the Evaluation of Educational Achievement (IEA) as part of the Trends in International Mathematics and Science Study (TIMSS) 2015 (Mullis et al. 2016a; 2016b), the intention of this brief is to undertake such an analysis and to answer questions such as whether students with grade level A in, for example, mathematics, perform better on average in TIMSS than students with a grade level B in mathematics.[1]

The analysis is carried out on students' test scores from TIMSS 2015 (year 8), their final school grades in mathematics and science in year 9, and their grades in national tests in mathematics and science in year 9. However, the associations between TIMSS scores and final grades and TIMSS scores and national test grades are similar. The size of the associations differs slightly, but the patterns are the same. Therefore, with the exception of Table 1, only the results of the final grades are presented.

School grades show students' accumulated knowledge at the end of the semester in relation to the knowledge requirements contained in the subject syllabus. Students receive a semester grade each semester from year 6 and a final grade in their last term in year 9. The grade scale goes from A to F, where A–E are passing grade levels and F indicates that the student has not passed the subject. There are specified knowledge requirements for grade levels A, C, and E, while the intermediate grade levels B

and D are given when the student has largely passed the requirements for a higher grade level. When calculating the total value of a student's final grades, the grade levels A, B, C, D, E, and F are coded numerically with the values 20.0, 17.5, 15.0, 12.5, 10.0, and 0.0, respectively.

National tests are given to students in year 3, 6, and 9 in some of the school subjects. The results from the national tests are transformed into the grades A–F, in the same way as the school grades.

With the purpose of this study and analytical methods in mind, we recoded both the final grades and national test grades so that A, B, C, D, E, and F are represented numerically by the values 6, 5, 4, 3, 2, and 1, respectively.

The content domains of the science part of TIMSS are biology, physics, earth sciences, and chemistry. Science subjects in Sweden include biology, physics, and chemistry. As the relationships between the final grades for these three subjects and the TIMSS results in science are very similar, we use in the analyses a weighting of the students' final grades in biology, physics, and chemistry. It gives us one final grade in science to analyze instead of three. For the same reason, we correspondingly analyze the associations between the national test grades and the TIMSS results in science weighed together instead of separately.

Personal identity numbers were collected during TIMSS 2015[2] enabling TIMSS scores to be merged with register data containing students' final grades and their results in national tests to allow for coherence analysis. We also carry out regression analysis in order to control for student gender, migration background, and number of home resources for learning. We are then able to investigate the coherency between TIMSS test scores and students' final grades with respect to the mentioned student background characteristics.

MAIN RESULTS

Clear differences in distribution of TIMSS scores between the grade levels

Figure 1 shows the distribution of TIMSS 2015 (year 8) scores in mathematics for each of the final grade levels (A–F)

in mathematics (year 9). Long bars indicate a large spread of students' results, while short bars indicate a small spread. The highest and lowest results are not reported because extreme results can contribute to a spread that does not reflect the group as a whole.

---

1. In this brief, "year" refers to year of schooling while "grade" refers to students' assessed level of knowledge at their last year in school and their national test results.
2. The national study center in Sweden collected and handled personal identity numbers according to Data Protection Legislation. They were collected as a national adaptation for Sweden only and exclusively available to the TIMSS national study center at the Swedish National Agency for Education.

Figure 1: Means and variations of TIMSS 2015 scores (year 8) in mathematics for the final grade levels in mathematics (year 9)
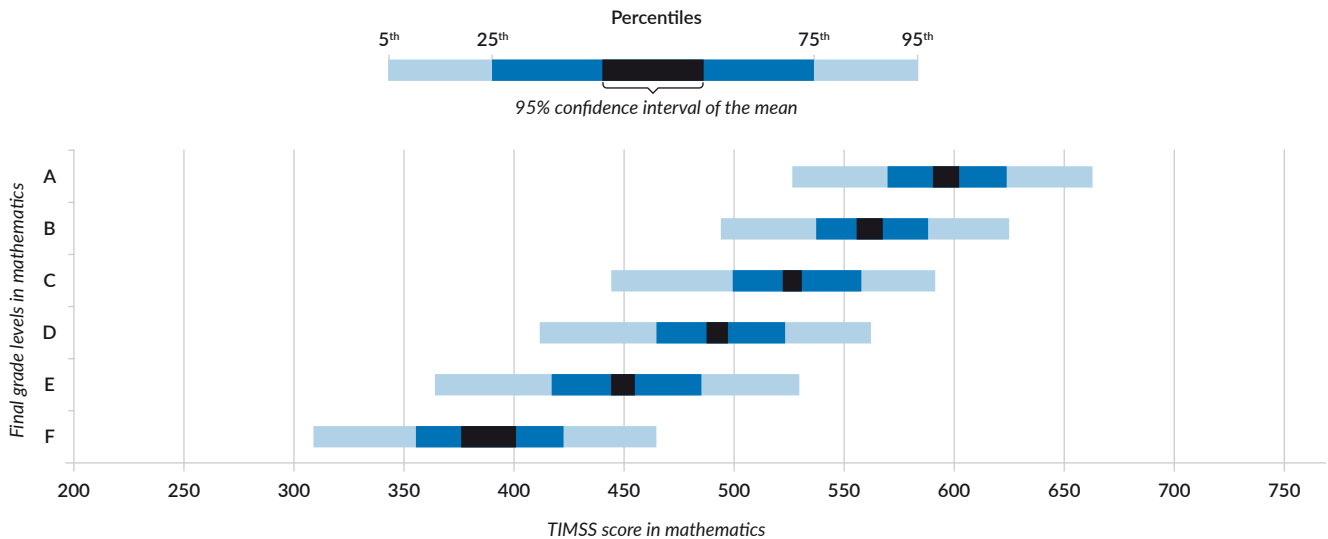

The black fields in the bars indicate the TIMSS mean scores at each grade level with associated 95 percent confidence intervals. The confidence intervals show the uncertainty in the estimates that results from TIMSS being a sample survey. Since the black fields do not overlap for the different grade levels, there is a statistically significant difference between the groups' mean points. The trend is that students with higher grade levels on average perform better in TIMSS mathematics. Furthermore, the black field is wider at grade level F, illustrating the larger uncertainty in the estimation of the mean at grade level F due to fewer students in that group. The dark blue fields, together with the black fields, cover the scores for half of the students as they constitute the score ranges for the 50 percent of students who perform closest to the median.

Furthermore, it is worth noting that the difference in mean scores between two adjacent grade levels is generally around 40 points. The difference, however, is larger between grade levels E and F.

Figure 2 shows the distribution of TIMSS 2015 (year 8) scores in science for each of the final grade levels (A–F) in science (year 9).

Figure 2: Means and variations of TIMSS 2015 scores (year 8) in science for the final grade levels in science (year 9)
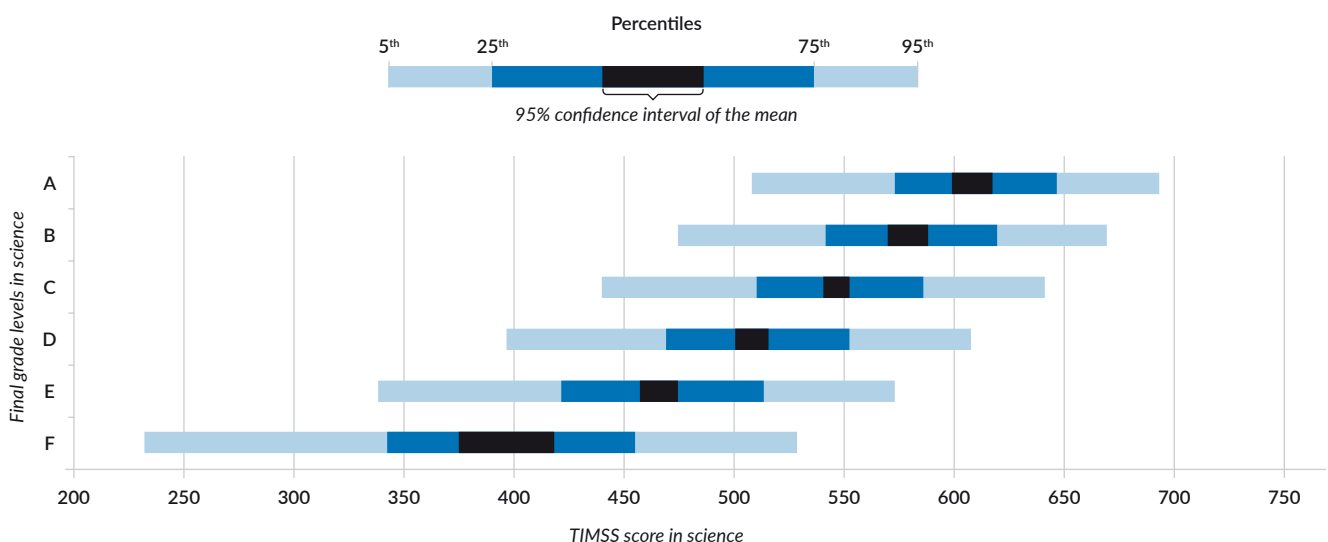
Compared to the distribution of the TIMSS scores in mathematics in Figure 1, the patterns are the same except that the distributions are offset by a few points at each grade level. Furthermore, the bars are longer which illustrates the larger spread of TIMSS results in science than in mathematics at each grade level, A–F.

We also note that the spread of TIMSS scores in science is larger among the students who received an F in science. This result may be due to the fact that the abilities of these students, in practice, vary widely—from having very large deficiencies in their knowledge to being close to a passing level. However, it may also indicate lower reliability of the measures—students more often do not take TIMSS-like low stake tests seriously. We do not see the same pattern in mathematics where the lengths of the bars are more evenly distributed over the grade levels.

Grades correlate moderately with TIMSS test scores

Pairwise correlations between students' final grades, their national test grades, and their TIMSS scores are given in Table 1. The correlations between final grades and TIMSS scores as well as national test grades and TIMSS scores

are rather moderate. For example, the correlation between students' final grades in mathematics and their TIMSS scores in mathematics is 0.76, and the correlation between students' final grades in science and their TIMSS scores in science is 0.64. Two plausible explanations are that the two measures partly cover different content and that the final grade is a measure based on other relevant information about the students' skills and knowledge development over a longer period. As pointed out in the previous section about lower reliability, the moderate correlations could also be due to the fact that ILSA results, of which TIMSS is one example, do not affect students' final grades and therefore may lower the motivation of some groups of students.

We further note that the correlation between students' final grades in mathematics and their final grades in science is stronger than the correlation between students' final grades in mathematics and their TIMSS scores in mathematics as well as the correlation between students' final grades in science and their TIMSS scores in science. The stronger correlation, 0.81, indicates that the Swedish context, or some general latent construct, is an important component in the strength of the associations between different performance measures.

Table 1: Correlations between final grades, national test grades, and TIMSS scores

		TIMSS (year 8)		Final grade (year 9)		National test grade (year 9)	
		Mathematics	Science	Mathematics	Science	Mathematics	Science
TIMSS (year 8)	Mathematics	1					
	Science	0.82	1				
Final grade (year 9)	Mathematics	0.76	0.66	1			
	Science	0.65	0.64	0.81	1		
National test grade (year 9)	Mathematics	0.77	0.67	0.92	0.75	1	
	Science	0.62	0.63	0.72	0.84	0.70	1

Regression analysis

Three linear regression models are fitted with TIMSS mathematics scores as the dependent variable. Model 1 includes only the student characteristics: sex (coded 1 if the student is male and 0 if the student is female), migration background 1 (coded 1 if the student is born in Sweden with both parents born abroad and 0 otherwise), migration background 2 (coded 1 if the student is born abroad and 0 otherwise), and home resources for learning (coded 1 if the student has many home resources and 0 otherwise) as explanatory variables. Model 2 includes only mathematics final grade as an explanatory

variable and model 3 includes the background variables and mathematics final grade as explanatory variables. Three corresponding linear regression models are fitted with TIMSS science scores as the dependent variable. Estimated regression coefficients are given in Table 2.

A cubic polynomial of the final grades was also included in model 2 and model 3 to account for curve linear associations between TIMSS scores and final grades. However, such associations turned out to be statistically non-significant.

Table 2: Estimated regression coefficients and their standard errors with TIMSS mathematics test scores and TIMSS science test scores, respectively, as dependent variables. The results are given for three regression models.

VARIABLES	Mathematics			Science		
	Only background variables	Only final grades	Both final grades and background variables	Only background variables	Only final grades	Both final grades and background variables
Intercept	**304.4** *(12.0)*	**448.3** *(2.5)*	**387.4** *(7.0)*	**293.7** *(12.0)*	**465.3** *(4.0)*	**356.9** *(9.4)*
Final grades		**38.3** *(1.1)*	**35.3** *(1.1)*		**38.0** *(1.5)*	**32.7** *(1.5)*
Sex	**10.5** *(3.0)*		**10.7** *(2.3)*	3.6 *(3.0)*		**14.2** *(2.8)*
Migration background 1	-8.2 *(6.3)*		**-13.9** *(4.7)*	**-24.5** *(7.5)*		**-31.1** *(6.2)*
Migration background 2	**-30.8** *(6.3)*		**-26.4** *(4.8)*	**-59.6** *(7.7)*		**-51.9** *(6.9)*
Home resources	**17.6** *(1.1)*		**5.7** *(0.6)*	**21.3** *(1.1)*		**10.6** *(0.9)*
Explained variance of TIMSS scores	**20%**	**57%**	**61%**	**25%**	**39%**	**49%**

Notes: *( ) standard errors are given within parentheses.*
*A **bold** coefficient means it significantly differs from 0.*

For the regression models with only final grades as an explanatory variable, 57 percent of the variance of the TIMSS scores in mathematics is explained by the final grades in mathematics and 39 percent of the variance of the TIMSS scores in science is explained by the final grades in science. By including student characteristics in the model, the explained variance increases by 4 percentage points with TIMSS mathematics scores as the dependent variable and by 10 percentage points with TIMSS science scores as the dependent variable. The relatively small increase of explained

variance indicates that a large part of the variation in TIMSS scores is explained by variation of final grades.

The relationship between TIMSS mathematics scores and mathematics final grades remains when we control for student characteristics—the small decrease from 38 points to 35 points is not statistically significant (Table 2). The corresponding effect of final grades in science on TIMSS science scores is a marginal, although significant, decrease, from 38 points to 33 points.

These small reductions of effects when controlling for student characteristics suggest that the abilities that make students with higher final grades outperform, on average, students with lower final grades in the TIMSS mathematics or science test, cannot be attributed to the different performance between males and females, between students with different home resources for learning, or between students with different migration background. There are likely other explanatory mechanisms, such as student interest in mathematics or science, or student reading ability.

Conversely, we also examined what happens to the relationship between TIMSS mathematics scores and student characteristics when we control for final grade in mathematics. If the TIMSS mathematics test accurately captures the Swedish curriculum for mathematics, we expect the effect of the background variables on TIMSS scores to disappear, or at least to be substantially reduced. The difference in TIMSS mathematics scores between students with lower and higher home resources decreases from 18 score points to 6 score points (Table 2). Although not completely controlled for by final grade in mathematics, the relationship is substantially reduced suggesting that the final grades in mathematics for these two student categories are highly related to the TIMSS mathematics results. The same conclusion holds for the relationships between TIMSS science scores and background variables when controlling for final grade in science.

However, the average differences in TIMSS mathematics scores between males and females still holds when controlling for mathematics final grades (Table 2). Thus, for a given grade level, males on average still have about 10 TIMSS mathematics score points more than females. Since the gender effect still remains after controlling for final grades, the difference in TIMSS scores between males and females is not controlled for by final grades. It therefore appears that whatever the mechanisms or abilities that make males on average outperform females in TIMSS mathematics are, they are not tested or considered in the national final grades. The average differences in TIMSS mathematics scores between students with different migration backgrounds also still holds when controlling for mathematics final grades:[3] for a given grade level, students born in Sweden with at least one parent born in Sweden have on average 14 mathematics score points more than students born in Sweden with both parents born outside of Sweden, and on average 26 mathematics score points more than students born outside of Sweden (Table 2). Equivalently to the conclusion of the remaining differences in TIMSS mathematics scores between males and females, the mechanisms or abilities that make students born in Sweden with at least one parent born in Sweden outperform, on average, students with other migration backgrounds in TIMSS mathematics, appears not to be tested or considered in the national final grades.

Similar conclusions can be drawn for corresponding analyses on the associations between TIMSS science scores, final science grades, and student characteristics (Table 2).

CONCLUSION

The international study, TIMSS, and the Swedish national grading system are important parts of the Swedish system for assessing students' knowledge. This report shows that both measures of students' knowledge have a clear coherence. However, it is important to emphasize that the international and national systems do not measure exactly the same thing. The correlation analyses that show moderately strong, positive correlations between the students' grades and their results in TIMSS confirm this.

---

3.   The only difference between the effects for a model with only background variables as covariates and a model with both background variables and final grades as covariates are due to sampling errors.

FURTHER READING

Skolverket. (2017). *TIMSS, TIMSS Advanced och betygen. Analys av sambandet mellan svenska betyg och de internationella TIMSS-studierna* [TIMSS, TIMSS Advanced and the grades. Analysis of the connection between Swedish grades and the international TIMSS studies]. Stockholm, Sweden: Skolverket. https://www.skolverket.se/getFile?file=3780

Wiberg, M., & Rolfsman, E. (2019). The association between science achievement measures in schools and TIMSS science achievements in Sweden. *International Journal of Science Education, 41*(16), 2218–2232. https://doi.org/10.1080/09500693.2019.1666217

Wiberg, M. (2019). The relationship between TIMSS mathematics achievements, grades, and national test scores. *Education Inquiry, 10*(4), 328–343. https://doi.org/10.1080/20004508.2019.1579626

REFERENCES

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in mathematics.* Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2015/international-results/

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016b). *TIMSS 2015 international results in science.* Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2015/international-results/

ABOUT THE AUTHORS

DR CHRISTIAN TALLBERG

Dr Christian Tallberg is a statistician at the Swedish National Agency for Education. Over the past 10 years he has worked with data from international large-scale assessments. He is also a co-writer to the Swedish national reports on PIRLS and TALIS. His main research interests are in statistical modelling of complex data structures. In addition to publications for the Swedish National Agency for Education, he has published in *The Journal of Mathematical Sociology* and *Social Networks.*

DR MARIA AXELSSON

Dr Maria Axelsson is Director of Education at the Swedish National Agency for Education. Since 2010 she has been working as the Swedish project manager for various international large-scale assessments, including TIMSS 2015 and 2019, as well as for various national evaluations. She is a graduate teacher and has previously worked at the Swedish Schools Inspectorate. She received her PhD in Economic History at Uppsala University, where financial history with gender perspective was her major research interest.

IEA COMPASS

ABOUT IEA

The International Association for the Evaluation of Educational Achievement, known as IEA, is an independent, international consortium of national research institutions and governmental agencies, with headquarters in Amsterdam. Its primary purpose is to conduct large-scale comparative studies of educational achievement with the aim of gaining more in-depth understanding of the effects of policies and practices within and across systems of education.

Thierry Rocher
IEA Chair

Dirk Hastedt
IEA Executive Director

Andrea Netten
Director of IEA Amsterdam

Gina Lamprell
IEA Publications Manager

Compass Editor
David Rutkowski
Indiana University

Follow us:

@iea_education

IEAResearchInEducation

IEA