The relationship between TIMSS mathematics achievements, grades, and national test scores

Marie Wiberg

Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden

ABSTRACT
The aim of this paper was to examine the relationship between Trends in Mathematics and Science Study (TIMSS) mathematics achievement and the two school achievement measures of grades and national test results in Sweden. A further aim was to examine the association of TIMSS mathematics achievement with different subgroups of students. The results show that there is a strong positive relationship between TIMSS mathematics achievement and national tests and obtained grades from both grade 6 and grade 9. Students from more educated homes performed better overall on TIMSS than those from less-educated homes, and the relationships between TIMSS and the school achievement measures were stronger for students from more educated home backgrounds. The school context and the students’ background had an impact on the students’ TIMSS result. The results have implications for how one should view the results from TIMSS as a measure of student mathematics achievement and thus how TIMSS results can be used in a national context.

KEYWORDS
TIMSS; mathematics achievement; correlations; multilevel analysis; national tests

Introduction
Results from Trends in Mathematics and Science Study (TIMSS) assessment in 2011 showed that Sweden had declining results in mathematics achievement (Mullis, Martin, Foy, & Arora, 2012). This pattern changed for TIMSS 2015, where Sweden showed increasing results in mathematics achievement (Mullis, Martin, Foy, & Hooper, 2016). An ongoing discussion in Sweden is whether TIMSS is measuring different or similar things as national tests in mathematics and school mathematics grades. Sollerman and Petterson (2016) performed an alignment study of the Swedish mathematics curriculum with the national tests in mathematics and TIMSS mathematics. The overall conclusions were that TIMSS is a relevant instrument to measure mathematics skills among Swedish students concerning the framework and the mathematics content in the items, although not all parts in the national curriculum were included in the assessment. Although TIMSS was concluded to be relevant in the Swedish context, the relationships with school grades and national tests were not completely clear.
Examining the relationship between international and national measures is important because if the relationship is strong, we can use information from TIMSS to support political decisions about a nation’s schooling. If the relationship between national and international assessments is weak and international assessments results are low for a nation, educational policy makers might decide to change the national curriculum or make other educational policy changes. Lipsy and Kijima (2016) made a systematic review of the impact of cross-national assessments, like TIMSS, on education policy outcomes. They concluded that these kind of assessments increase both the capacity and motivation of policymakers to implement improvements in education. In their study, a US delegate said that the TIMSS and PISA (Programme for International Student Assessment) results helped to justify the reforms of the No Child Left Behind Act. Yariv (2018) concluded that policymakers’ discourse regarding mathematics education transformed based on the TIMSS 1999 results and to improve Israel’s ranking, the middle school mathematics curriculum was aligned with the TIMSS curriculum. The declining results in international assessments in Japan were used politically to abandon creativity under a relaxed curriculum, and instead implement a more traditional and academic rigour curriculum (Takayama, 2008). Educational reforms as a direct response to results of international assessments can be observed in other countries as well (see e.g. Grek, 2009; Lingard & Grek, 2006; Martens, Nagel, & Windzio, 2010). International assessments and national assessments might not always show the same pattern. For example, in Canada international assessment results and provincial assessment results often differ (Cartwright, Lalancette, Mussio, & Xing, 2003). There have also been inconsistent results from TIMSS and PISA for some countries (Wu, 2009). In Poland, national assessment trends have not always been consistent with trends in PISA (Szaneniec, Grudniewska, Kondratek, Kulon, & Pokropek, 2013). There is therefore an interest in examining the relationship between international and national assessments within each country.

It is well-known that students, who take national and international assessments, are not a homogenous group in terms of their home background. Many different researchers have concluded that students from homes with a high socioeconomic background are more likely to perform well in school and to have better academic performance than their peers with lower socioeconomic background (e.g. Erberber, Stephens, Mamedova, Ferguson, & Kroeger, 2015; Hanushek & Luque, 2003; Jurdak, 2014; Sirin, 2005; Wikström, 2005). This association is universal across nations, school subjects (e.g. science and mathematics), and grades (e.g. from primary to secondary education) (Erberber et al., 2015; OECD, 2011). Akyuz and Berberoglu (2010) examined TIMSS data from nine European countries and found that home educational resources were significantly related to mathematics achievement in eight of the countries. Teodorovic (2011) concluded, after reviewing school effectiveness research that student-level variables are very important in determining student achievement in rich countries but are less pronounced in poor and developing countries. Kaleli-Yılmaz and Hanci (2016) examined TIMSS 2011 mathematics achievement in Turkey in connection with different student variables, including gender, school grades, parents’ educational level, and cognitive domains. They did not find any gender effect, but they did find a relationship with TIMSS and school grades and parents’ educational level. In their study, if the student’s mother had a high educational level the student had on average higher TIMSS achievement than if the mother had a lower educational level.
The average score on TIMSS was also higher when the mothers’ educational level was used to divide the students compared with if the students’ fathers’ educational level was used.

Caponera and Losito (2016) examined contextual factors and student achievement in all participating countries in TIMSS 2011 and concluded that high socioeconomic status has a significant and positive effect on student achievement and that students from advantaged schools perform better. For Sweden, especially the student’s gender and socioeconomic status and the average student socioeconomic status in the school were significant in their study. In Hastedt (2016), it was concluded that an immigrant background can affect student achievements either positively or negatively depending on the background of the immigrants. From the results of the referred studies, we believe it is of interest to examine TIMSS mathematics achievement in connection to students’ home background variables as well as the school context in terms of the composition of students within the schools. Summing up, this study aims to examine the relationship between TIMSS mathematics achievement and the two national school achievement measures of grades and national test results in Sweden. A further aim is to examine the association of TIMSS mathematics achievement with different subgroups of students.

In order to analyse TIMSS mathematics achievement and its association with different student groups, several country-specific studies have used multilevel analyses where one takes into account both student variables and school context variables. However, these studies yield different conclusions depending on which countries are examined. For example, Ghagar, Othman, and Mohammadpour (2011) examined Malaysian and Singaporean TIMSS mathematics achievement with multilevel analyses and concluded that there were student variables that were significant in both countries and that there were school differences in Malaysia, while the differences between classrooms were more evident in Singapore. Harmouch, Khraibani, and Atrissi (2017) modelled Lebanese and Singaporean TIMSS mathematics achievement and concluded that both student variables and school context were relevant, but which variables that were significant differed between countries. Wiberg, Rolfsman, and Laukaityte (2013) used multilevel analyses to model Swedish TIMSS mathematics achievement in 2003, 2007, and 2011. They concluded that students’ home background variables as well as the school’s average socioeconomic background were significant, while other school-level factors were not. Thus, besides examining the association between student variables and TIMSS achievement, the composition of students in the schools might also be of interest when modelling TIMSS mathematics achievement in Sweden.

Most TIMSS studies have only used material from the TIMSS international database (e.g. Caponera & Losito, 2016; Kaleli-Yılmaz & Hanci, 2016; Wiberg & Rolfsman, 2013; Wiberg, Rolfsman & Laukaityte, 2013; Drent, Meelissen, and van der Kleij (2013) for a review of TIMSS studies until TIMSS 2011). This study is unique because it combines TIMSS mathematics achievements with other school achievement measures and register data. Thus, the obtained results are important in Sweden as it may affect how TIMSS results are used and which value they are given in comparison to grades and national test results. Although the study is conducted within the national context of Sweden, the results might be of interest in other countries because many countries participate in the international TIMSS assessment, most countries use subject grades, and many countries have national tests. We also believe that this study is important in a wider context because we
have not been able to find any similar studies with TIMSS data. The closest studies we have found are studies that link different large-scale assessments, such as the linking between TIMSS and the US National Assessment of Educational Progress (NAEP) scale (e.g., Johnson, 1998; Hambleton, Sireci, & Smith, 2009; Lim & Sireci, 2017). The focus in those studies was on how students from other countries would likely perform on the NAEP by linking the NAEP scale to TIMSS and PISA both for a particular year and over time. For example, in Lim and Sireci (2017) an equipercentile equating between NAEP and TIMSS was carried out, and Hambleton et al. (2009) compared the students’ performance on NAEP through the link between NAEP and TIMSS and PISA. A major finding was that the standards for NAEP are set high (especially on the highest level) but not too high in an international context. Those studies are thus very different from this study because they did not examine the individual link between an individual’s different international and national achievement measures but rather were studies on the overall link between two different achievement measures.

Although we have not been able to find any similar TIMSS studies, there are studies that have examined the relationship with the large-scale Scholastic Assessment Test (SAT) scores and high school grades and socioeconomic factors (e.g. Zwick & Green, 2007; Zwick & Himelfarb, 2011), including the role of student ethnicity and first language (Zwick & Sklar, 2005). In these studies, it is evident that the SAT and grades have similar connections to socioeconomic factors if one takes into account the school context. Students’ socioeconomic status correlates highly with both SAT scores and grades, and students who have a different first language than English are disadvantaged on the SAT. It is important to note that although the student background data were self-reported, these results are interesting because even if TIMSS and the SAT are not the same kind of measure they share similar features. They are, for example, freer from the curriculum than tests given in schools or national tests, which are based completely on the curriculum. Recently, some preliminary analyses were conducted by the National Agency for Education (2017) in Sweden with a focus on TIMSS and grade 9 in mathematics and science. This study differs from that report because our focus is on the mathematics results and thus we also include the student grades and national test results from grade 6 and control for somewhat different student background variables. Based on our aims and the literature review, we thus focus on the following research questions:

1. How is student achievement in school in terms of grades and national test results associated with TIMSS mathematics achievement in Sweden?
2. How are student home background variables associated with TIMSS mathematics achievement in Sweden?
3. How is the school context in terms of average level of student background variables associated with TIMSS mathematics achievement in Sweden and can the school context explain some of the variation in TIMSS mathematics in Sweden?

We hypothesised that there is a high degree of association between TIMSS and the national school achievement measures, but we expected the national achievement measures to be more highly correlated with each other because they follow the same
curriculum. From previous research of Swedish TIMSS data (e.g. Caponera & Losito, 2016; Wiberg et al., 2013), we expected student-level variables as well as the school context in terms of the composition of students at the school level to be associated with TIMSS mathematics achievement. We however expect that only a small part of the variance can be explained by the school context.

In the next section, the method used is described, including the participants, the instruments used, and the statistical analyses conducted. The third section contains the results, and the paper ends with a discussion section where some concluding remarks and some practical implications are given.

Method

Participants

Students’ mathematics achievement results from TIMSS 2015 8th grade in Sweden were used (IEA, 2017). The students in TIMSS 2015 were randomly selected in a two-stage procedure so they are representative of all the grade 8 students in Sweden. In TIMSS 2015, the students came from 150 different schools, and a total of 4,090 students participated in grade 8 from Sweden. In the used sample, there were 52% boys, 20% who were nonnative born or who had nonnative-born parents, 46% who had mothers with higher than high school education, and 37% who came from homes with more than 100 books. The response rate in Sweden tends to be generally high, and only about 5% did not respond to the TIMSS items. The nonresponsive students were students who were out of school the day TIMSS was given at their school. The instruction to the students who were part of the TIMSS assessment was that the test is important. However, because the results are not used for grading the student it is a low-stakes test. The TIMSS 2015 assessment offered a unique possibility to carefully study TIMSS in Sweden because social security numbers were collected from all students and this made it possible to connect the information in TIMSS with Swedish register data containing information about national test results, school grades, immigrant status, and parental education level.

Instruments

National tests

The Swedish national tests are given in grades 3, 6, and 9 in different core subjects and comprised different subtests. The starting point for the tests is the curriculum and the syllabus. The aim of the national tests is to support an equal and fair grading process and to provide information about how the knowledge demands are fulfilled on a school level and on a national level (National Agency for Education, 2016). The national tests are not final tests, but are rather one out of many other tests that help the teacher to provide a fair grade to the student. The mathematics national test aims to measure the students’ different mathematics abilities through the use of four (grade 9) or five (grade 6) subtests. One subtest is oral and is given in groups of 3 or 4 persons and takes about 20–30 (grade 9) or 60–80 minutes (grade 6) to complete. The three (grade 9) or four (grade 6) written subtests are individual and given on specific national tests days. The
written subtests contain items that require responses in the form of multiple choice, short answers, or complete solutions. Each written subtest takes about 80–100 minutes (grade 9) or 40–80 (grade 6) minutes to complete. The national tests are designed to give the teachers a tool so they can provide a fair grade to each of the students regardless whether they are a low-performing or a high-performing student. This means that the national tests are high-stakes tests for the students. In this study, we had access to national test results in mathematics and other subjects from grade 6 and grade 9.

The Swedish grading system
The purpose of subject grades is to measure students’ subject knowledge. The current grading system in Sweden is criterion-referenced and is an evaluation of a student’s knowledge in comparison to specific knowledge criteria. The grading scale comprises six steps, A–F, where F is fail and A–E are different levels of passing with A as the top grade. For each grade step, there are rubrics that describe what a student needs to know in order to obtain a particular grade. The grades are determined by the teacher at the end of a course or at the end of a school semester. If there is too little information about a student, then no grade is given to that student. The letter grades also correspond to the following numeric grading scale: A = 20, B = 17.5, C = 15, D = 12.5, E = 10, F = 0. The numeric grading scale allows us to get a joint metric of the sum of the 16 subject grades a student receives in grade 9. This overall grade value is referred to as the merit value and ranges from 0 to 320 because the best-performing student getting all As would obtain $16 \times 20 = 320$ (National Agency for Education, 2017). Because the merit value contains 16 grades, it has a larger variability because many teachers set the different grades. The value could, however, be viewed as a measure of how well the student performs overall in school. In this study, we had access to the students’ grades in mathematics from grade 6 and grade 9 as well as their merit values from grade 9.

Grades vs. national tests
Grades and national tests are different in several respects. The national test results stem from a single test or a few single tests given on a single occasion that test many parts of the curriculum, but not the whole curriculum. Grades on the other hand are an overall judgment of the student that typically builds on several assessments on several occasions and covers the whole curriculum.

TIMSS 2015
TIMSS is an international assessment of mathematics and science given every four years since 1995, and the respondents are students from grade 4 and 8. The idea is to examine trends in student achievement together with some contextual data. Fifty-seven countries participated in TIMSS 2015, and the assessment was constructed on frameworks developed by the participating countries for each curriculum area and for each grade. Most items are constructed to assess students’ application and reasoning skills (Mullis et al., 2016). In TIMSS, the students’ achievements are summarised with five plausible values. In this study, the five plausible values were used for mathematics achievement from grade 8 for Swedish students.
**TIMSS 2015 levels**

A number of benchmarks define the achievement levels in TIMSS 2015 that the student can be categorised within. These levels are defined from the TIMSS achievement values and include the elementary level (400–474 points), average level (475–549 points), high level (550–624 points), and advanced level (625 or more points) (Mullis et al., 2016). In this paper, we will examine these levels in connection to the students’ school grades.

**Comparisons between TIMSS, grades, and national tests**

We used data on students’ grades in mathematics and their national test results in mathematics from grade 6. At the time the international database from TIMSS 2015 was released, the students had received their final grades in grade 9 as well as had taken national tests in grade 9 in the core subjects that are assessed in TIMSS. We used national mathematics test results from grade 9 and information about the students’ mathematics grade and a summary of 16 different subject grades from grade 9 (merit value). We are aware that the achievement measures used in this study were not measured at the same time that TIMSS 2015 was given to grade 8 students, and we used school achievement measures from grade 6 and grade 9. To use information from different grades gives us information about whether the observations are consistent over time or not. We are aware that the students in grade 9 might have matured and that they had learned more between grade 8 and grade 9. However, we think they are relevant to use because the national tests share similar features to the international TIMSS assessment. They both are given on certain test days with certain subtests, and the results are later used in the public discussion about the students’ achievements. The national tests are also used to ensure that the grades from one school are not inflated. Thus, the grades from grade 9 and national tests from grade 9 should be somewhat close to each other. We thus examined the connection between different school grades and national test results and TIMSS mathematics achievement, but we did not use the national school achievement measures from grade 9 to try to explain the TIMSS mathematics achievement. In summary, we compared the students’ TIMSS mathematics achievement results with their mathematics grades from grade 6 and grade 9, with their results from national tests in mathematics from grade 6 and 9, and with the their merit values from grade 9.

**Statistical analysis**

We used descriptive statistics such as average scores, standard errors, and correlations between TIMSS 2015 mathematics achievement and national test results and the students’ grades from grade 6 and 9 to answer the first research question. All five mathematics plausible values from TIMSS 2015 were used in the analyses as a measure of the students’ mathematics achievement in line with the recommendations on how to use plausible values given in Laukaityte and Wiberg (2017). Student weights were used on the student level when conducting the analyses. For the statistical analyses we used IEA IDB analyser 4.0.12 (2018), SPSS 24.0 and HLM.

To answer the second research question, we examined the association between TIMSS mathematics achievement in Sweden and the students’ home background because previous studies have shown that a student’s socioeconomic status might be of importance when examining students’ achievements. There is no universal definition
of socioeconomic status, but basic dimensions include parents’ education level, parents’ occupation, and family income. Unfortunately, we did not have access to family income data and the parents’ occupation contained too much missing information. We did, however, have access to the parents’ educational level. Because the mothers’ educational level contained fewer missing observations, and inspired by the results of Kaleli-Yılmaz and Hanci (2016) as described in the introduction, the mothers’ educational level was used instead of the fathers’ educational level. Note that we also reran the analysis using the fathers’ educational level and got similar results. We chose not to use the TIMSS home educational resource index, which is based on the following TIMSS variables: number of books at home, possession of an Internet connection, if the student has their own room, and the parents’ highest educational level. One reason was that we thought it was more interesting to examine the influence of selected variables instead of an overall index. Another reason was that most Swedish students have an Internet connection (99%) and their own room (93%). Not using home possessions differs from the study by Caponera and Losito (2016). Although sex is not part of defining socioeconomic background, it was shown to be significant for Sweden in the study by Caponera and Losito (2016) and it is of interest in Sweden. For example, the national test results in mathematics from grade 6 and grade 9 do not show any large differences between boys and girls (National Agency of Education, 2016), but a previous study has shown that girls tend to get better grades than boys in relation to their performance on national tests (Nycander, 2006). Although girls tend to get better grades at the end of compulsory school in all subjects except sports/athletics, the difference is smallest in mathematics, physics, and technology. We thus did not expect it to have a large impact on the relationships examined here. Migration, in terms of whether students or their parents were born in Sweden or not, was examined because it is well known that it can affect mathematics achievement in both a positive and a negative way depending on the background of the immigrants, as seen in Hastedt (2016). In summary, student background variables were chosen from the students’ questionnaires based on previous studies (e.g. Caponera & Losito, 2016; Ilie & Lietz, 2010; Wiberg & Rolfsman; 2013) and the availability of data in TIMSS and in the register we had access to.

Linear regressions were used to model TIMSS mathematics achievement with the student home background variables in order to answer the second research question. They were also used for comparison when examining the third research question connected to the school context, which was primarily examined with multilevel analyses. Multilevel analyses (Snijders & Bosker, 2012) are often used to analyse large-scale assessments, and such analyses take into consideration the two-stage sampling procedure of TIMSS. Because such analyses allow one to enter variables on several levels, they gives a deeper understanding of the relationship between TIMSS mathematics achievement and the school achievement measures, the student-level variables, and the average school-level student variables.

The students’ background variables were recoded to make sense in the analyses. The students’ sex was coded as 0 for girls and 1 for boys, and the migration variable was coded as 0 (labelled Swe) if the student was native born by native parents and coded 1 (labelled Imm) otherwise. The students’ mothers’ educational-level Med was defined as 1 if the mother had at least one year of education after high school education (labelled Hmed) and coded 0 otherwise (labelled Lmed). The variable book was defined as 1 if the
student had more than 100 books at home (labelled *Hbook*) and coded 0 otherwise (labelled *Lbook*). The students’ letter grades (A–F) were converted into the numeric grading scale (20–0) as described above in the grading system subsection. All of the home background variables examined in the later linear regressions and multilevel analyses were significant on at least the 0.05 level. Missing data were in general low in the students’ background variables, ranging from 0.3% (sex) to 7% (national test results from grade 6), thus listwise deletion was used to exclude missing data (Tabachnik & Fidell, 2007). The missing data in national tests were due to students who were absent on the day of testing. We are aware that it is theoretically better to impute missing data, but the choice to use listwise deletion was made because reasonably few cases were deleted. Note that we only used grade 6 school achievement measures as independent variables in the linear regressions and multilevel analyses because it would be improper to use grade 9 measures to explain TIMSS mathematics achievements obtained in grade 8.

**Results**

The results in Table 1–3 are related to the first research question. In Table 1, the average scores and standard errors on TIMSS 2015 in mathematics for students receiving different grade levels in grade 6 and grade 9 are displayed. From Table 1 it is evident that regardless of year, the higher the grades the students had, the higher the student’s average TIMSS achievement. Although this was true in both grade 6 and grade 9, the TIMSS results were in general higher for the different grade groups in grade 9. This is not surprising because there are more criteria that need to be met by the end of grade 9 in comparison to grade 6. Thus, this means that a smaller proportion of students might not have managed to get the highest mathematics grade in grade 9 (10%) in comparison to grade 6 (16%) because there are fewer goals to achieve in grade 6. This was similar for the students scoring highest on the national test for grade 9 (6%) compared to grade 6 (20%).

Next, we examined the proportions of students who received each grade in grade 6 and grade 9 distributed on the TIMSS achievement levels (Table 2). By merging the two last TIMSS categories of 550–624 points and 625 or more points, we could conduct chi-square tests that showed that grades and TIMSS achievement levels cannot be seen as independent either in grade 6 ($X^2(15) = 442$, $p < 0.001$) or in grade 9 ($X^2(15) = 594$, $p < 0.001$). Noticeable is the consistency of the overall results regardless of whether we

---

**Table 1.** The proportions of students (%) in grades 6 and 9 receiving grades A–F on the national mathematics assessments and their respective mean scores (M) and standard errors (SE) on the TIMSS mathematics assessment.

| Grade | Mathematics grade | National test in mathematics |
|-------|------------------|---------------------------- |
|       | grade 6          | grade 9                    |
|       | %                | M (SE)  | %                | M (SE)  |
| A     | 16               | 580 (2.8)| 10               | 596 (3.2)| 20               | 575 (2.9)| 6               | 607 (3.8)|
| B     | 18               | 543 (3.4)| 13               | 562 (3.2)| 16               | 535 (2.9)| 10              | 578 (3.1)|
| C     | 20               | 508 (2.8)| 20               | 526 (2.4)| 16               | 507 (3.3)| 20              | 542 (2.6)|
| D     | 20               | 476 (3.0)| 20               | 492 (2.4)| 21               | 473 (3.3)| 22              | 498 (2.5)|
| E     | 17               | 442 (3.0)| 29               | 449 (3.0)| 15               | 443 (3.0)| 28              | 458 (2.4)|
| F     | 5                | 404 (6.0)| 4                | 388 (6.3)| 5                | 413 (5.4)| 7               | 405 (6.0)|
|       | 3                | 475 (30.8)| 3               | 463 (9.1)| 2                | 458 (15.9)| 6               | 460 (8.6)|

---

M. WIBERG
use data from grade 6 or grade 9. It is also interesting to note that no student with a grade of A got fewer than 400 points on the TIMSS achievement scale. Likewise, no student with a grade lower than C reached the TIMSS advanced level.

The overall correlations, which were used to examine the relationships between grades, national tests, and TIMSS achievement can be seen in Table 3 in general and in Table 4 for different subgroups of students. The correlations between grades and national tests for both grade 6 (0.89) and grade 9 (0.86) were strong and positive. This indicates that the national test helps the teachers in their grading exactly as it is supposed to do. Recall that the aim of the national tests is to support an equal and fair grading process and to give information about how the knowledge demands are fulfilled on a school level and on a national level. From Table 3 it is evident that the highest correlation with TIMSS was for mathematics grade 9 (.74), followed by national test grade 9 (.73). This is not surprising if we believe that these achievement measurements measure similar things and because these were collected only one year after (grades and national tests in grade 9) or two years before (grades and national tests in grade 6) the TIMSS assessment was conducted. The correlations are quite high and in line with our expectations, although we thought they would be closer to the correlation between grades and national test results. It is interesting to note that the correlation between merit value and TIMSS was quite high, even though the merit value is a combined measure of 16 subject grades in grade 9. The merit value indicates what is usually seen in school studies – that students who perform well in one subject area also tend to perform well in other subject areas.

Table 2. Proportions of students within each of the grades in grade 6 and grade 9 depending on TIMSS achievement levels.

| TIMSS | A   | B   | C   | D   | E   | F   |
|-------|-----|-----|-----|-----|-----|-----|
| Grade 6 |     |     |     |     |     |     |
| (625 ≤) | 16  | 3   | 0   | 0   | 0   | 0   |
| (551,624) | 60  | 44  | 21  | 5   | 1   | 0   |
| (476,550) | 22  | 45  | 54  | 49  | 25  | 82  |
| (401,475) | 1   | 8   | 22  | 38  | 54  | 45  |
| ≤400   | 0   | 1   | 3   | 7   | 19  | 46  |
| Grade 9 |     |     |     |     |     |     |
| (625 ≤) | 24  | 5   | 1   | 0   | 0   | 0   |
| (551,624) | 63  | 59  | 32  | 9   | 2   | 0   |
| (476,550) | 12  | 34  | 56  | 58  | 30  | 3   |
| (401,475) | 0   | 2   | 11  | 29  | 52  | 38  |
| ≤400   | 0   | 0   | 1   | 0   | 16  | 59  |

Table 3. Correlations between TIMSS mathematics achievement and the different school measures of mathematics achievement.

| Variables | TIMSS | G6 | NT6 | G9 | NT9 |
|-----------|-------|----|-----|----|-----|
| G6        | .68   |    |     |    |     |
| NT6       | .67   | .89|     |    |     |
| G9        | .74   | .69| .70 |    |     |
| NT9       | .73   | .66| .65 | .86| .73 |
| Mvalue G9 | .66   | .61| .59 | .80| .73 |

G6 = Mathematics grade 6, G9 = Mathematics grade 9, NT = National test in Mathematics, and Mvalue G9 = Merit value in grade 9.
The results in Tables 4 and 5 were used to answer the second research question. When examining the sub correlations of the measures in Table 3, no significant differences in the relationships were found between girls and boys, as seen in Table 4. If the student was nonnative or had nonnative parents only gave a significant correlation difference in the merit value. This is not surprising because the merit value contains all subject areas and thus it is possible that language issues will have a stronger effect on that measure. The relationship was significantly different for most achievement measures if the mother had a high educational level or if the student lived in a home with more than 100 books. The exceptions were for national test score in grade 6 and students from homes with high number of books and the merit value. The latter result is probably due to the diversity among school subjects, which the merit value captures.

Correlation can only give us a measure of the linear relationship between two variables. To answer the second and third research questions, and to be able to examine the impact of several variables on both student and school levels, we used linear regressions and multilevel analyses. In the left part of Table 5 linear regressions and

| Variables | G6 | G6 | G6 | NT6 | NT6 | NT6 |
|-----------|----|----|----|-----|-----|-----|
| Intercept | 501.27 (2.09) | 502.26 (1.88) | 502.80 (1.74) | 503.77 (1.94) | 504.67 (1.74) | 504.46 (1.59) |
| Achieve* | 9.75 (0.30) | 9.06 (0.29) | 9.04 (0.29) | 9.24 (0.31) | 8.51 (0.30) | 8.47 (0.30) |
| Sex | 10.78 (1.96) | 10.87 (1.97) | 10.45 (2.16) | 10.45 (2.16) | | |
| Imm | -6.33 (2.82) | -5.58 (2.73) | -8.19 (3.05) | -5.78 (2.96) | | |
| Book | 16.96 (1.81) | 15.88 (1.84) | 16.46 (1.92) | 15.52 (1.94) | | |
| Med | 10.35 (1.86) | 9.96 (1.88) | 11.38 (2.00) | 10.94 (2.05) | | |
| School | | | | | | |
| Imm_A | -21.72 (9.35) | | | | | |
| Book_A | 57.93 (10.44) | 36.57 (9.59) | | | | |
| R² | 0.45 | 0.48 | 0.48 | 0.43 | 0.46 | 0.46 |

G6 = Mathematics grade 6, NT6 = National test in mathematics in grade 6. *Achieve = Achievement measure used varies and the achievement measure used is given as the name of each column. Sex = 1 if boy, 0 if girl. Imm = 1 if student or student’s parents not born in Sweden. Med = 1 if student’s mother has higher education than high school education, 0 otherwise. Book = 1 if student’s home has more than 100 books, Imm_A = aggregated Imm, Book_A = aggregated Book. R² = Explained variance.

Table 5. Regression and multilevel analyses with TIMSS mathematics achievement as the dependent variable and student background variables and different types of grades as the independent variables.
multilevel analyses are displayed for the subject grade 6 achievement measures and likewise in the right part of Table 5 these are given for the grade 6 national test achievement measure. Note that we examined all possible average student background variables on the school level, but only the significant school-level variables are shown in Table 5. The analyses indicated that on average being a nonnative student meant that TIMSS mathematics achievement was on average lower and being a boy meant on average a higher TIMSS mathematics achievement. A home with many books and a mother with a higher education were associated with higher TIMSS achievement in general. In addition, regardless of which achievement measure was used, being in a school with students with many books at home gave on average higher TIMSS achievement. In contrast, the aggregated immigration variable was only significant for national tests data. Students who received a higher grade in the national achievement measures had on average higher TIMSS mathematics achievement.

The multilevel analyses were similar to the linear regressions with student home background in terms of proportion of explained variance, which ranged from 0.46 (national tests) to 0.48 (grade 6). The explained variance with only the achievement measure and no other school background measure was a bit lower at 0.43 (national test) and 0.45 (subject grades). The intraclass correlation coefficients (ICCs) in the multilevel analyses were 0.11 (national test) and 0.13 (grade 6). Thus, the school context seemed to only have a small impact on the students’ results.

Summarising the findings from the perspective of the posed research questions, there seems to be a quite high association between TIMSS mathematics achievement and national school achievement in Sweden, although the association is lower than between the two national achievement measures. Further, there appears to be a strong association with TIMSS mathematics achievement and students’ home background variables when it comes to immigration and socioeconomic status in terms of books at home and parental educational level, while there is no strong association with gender. Finally, the school context in terms of the average student background seems to only be able to explain a small amount of the variance.

**Discussion**

The aim of this paper was to examine the relationship between grades, national test scores, and TIMSS mathematics achievement as well as to examine the association between TIMSS mathematics achievement and different subgroups of students. The overall result is that the relationship with TIMSS, national tests, and grades is strong in Sweden regardless of whether we examine the relationships from before TIMSS (grade 6) or after TIMSS was conducted (grade 9). A somewhat weaker relationship was observed if the joint grade from 16 subjects (merit value) was used as the achievement measure. This is not surprising because this measure contains greater variability because many different teachers are part of grading the students in the 16 subjects that the merit value is based upon. As expected and in line with, e.g. Nycander (2006), the gender differences were small. The educational background of the students’ mother and if the student came from a home with many books did, however, have a clear influence on the students’ academic performance. These results are in line with the results from a wide range of previous studies (e.g. Chiu & Xihua, 2008; Erberber et al., 2015;
We used multilevel analyses in order to take into account the hierarchical structure of TIMSS and its two-stage sampling design, where the probability of selecting a sample unit is unequal (Kyriakides & Charalambous, 2005). An advantage of such analyses is the possibility to examine whether the average level of school achievement influences the results of students’ achievement and to control for both student and school factors in the analyses. Although the aggregated number of books variable was significant when using grade 6 and national test scores in the analyses, this was not the case for the aggregated migration variable, which was only significant when using national test scores as the achievement measure. The results from the multilevel analyses were compared with the results from the linear regressions, and it was concluded that school-level variables only account for a small amount of the variance. As expected, there were only a few variables that were significant on the school level in Sweden in TIMSS 2015, which is in line with previous studies based on TIMSS 2003, TIMSS 2007, and TIMSS 2011 achievement in mathematics and science in Sweden (e.g. Rolfsman, Wiberg, & Laukaityte, 2013; Wiberg et al., 2013). The quite low ICC values are similar to other studies that have used international assessments. Because Sweden is a high-income country, it is also unlikely that the school has a large effect according to the Heyneman-Loxley effect, which states that the quality of schools has a greater impact on achievement in low-income countries than it does in high-income countries. It is also true that the effects on achievement of the student family context tend to be stronger in higher-income countries (Ilie & Lietz, 2010), something that could be seen in the multilevel analyses presented here.

This study also shared some resemblance with studies examining SAT scores (a large-scale assessment) and their correlation with high school grades and socioeconomic factors. The conclusion that TIMSS results were lower for nonnative students is in line with the results in the Zwick and Green (2007) SAT study that found that students with a first language other than English were disadvantaged on the SAT. A conclusion shared with other studies is that in everyday school life a significant challenge is how to help students from lower socioeconomic background homes and/or are nonnative students succeed in all subjects in school.

The uniqueness of this study was that it allowed for connecting TIMSS mathematics achievement with the individual students’ grades and results on national tests in Sweden. The use of grade 6 measures and background information among the students can also help in explaining the Swedish students’ TIMSS mathematics achievement. The strong positive relationship found between TIMSS mathematics achievement and school grades and national tests in mathematics in both grade 6 and grade 9 implies that Sweden can use TIMSS as a trend indicator for how the students perform over time within the schools and thus use the results more strategically. Although this study was set in a Swedish context, the results are important for other countries that take parts in TIMSS. It is likely that the results would be similar for other countries, and thus it would be interesting to repeat the conducted analyses in other countries.
Note

1. The SAT is a standardised test used for college admissions in the US, and thus it focuses on skills that matter for college readiness and success according to recent research. The test consists of a reading test, a writing and language test, and a mathematics test. There is also an optional essay that some colleges require.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Swedish Research Council [2015-02160].

ORCID

Marie Wiberg http://orcid.org/0000-0001-5549-8262

Notes on contributor

Marie Wiberg is professor in statistics with specialty in psychometrics at Umeå University, Sweden. Her research focus on large-scale assessments and standardized achievement tests. Of special interest are test equating methods, how to score tests and other research problems connected to large-scale assessments and standardized achievement tests.

References

Akyuz, G., & Berberoglu, G. (2010). Teacher and classroom characteristics and their relations to mathematics achievement of the student in the TIMSS. New Horizon in Education, 58(1), 77–95.

Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. Large-Scale Assessments in Education, 4(12). doi:10.1186/s40536-016-0030-6

Cartwright, F., Lalancette, D., Mussio, J., & Xing, D. (2003). Linking provincial student assessments with national and international assessments. Ottawa: Culture, Tourism and the Centre for Education Statistics, Statistics Canada. Retrieved from http://www.publications.gc.ca/Collection/Statcan/81-595-MIE/81-595-MIE2003005.pdf

Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of 41 countries. Learning and Instruction, 18(4), 321–336.

Drent, M., Meelissen, M. R. M., & van der Kleij, F. M. (2013). The contribution of TIMSS to the link between school and classroom factors and student achievement. Journal of Curriculum Studies, 45(2), 198–224.

Erberber, E., Stephens, M., Mamedova, S., Ferguson, S., & Kroeger, T. (2015). Socioeconomically disadvantaged students who are academically successful: Examining academic resilience cross-nationally (No. 5). IEA’S Policy Brief Series. Amsterdam, The Netherlands: IEA. Retrieved from http://www.iea.nl/policy_briefs.html

Ghagar, M. N., Othman, R., & Mohammadpour, E. (2011). Multilevel analysis of achievement in mathematics of Malaysian and Singaporean students. Journal of Educational Psychology and Counseling, 2, 285–304.

Grek, S. (2009). Governing by numbers: The PISA ‘effect’ in Europe. Journal of Education Policy, 24(1), 23–37.
Hambleton, R. K., Sireci, S. G., & Smith, Z. R. (2009). How do other countries measure up to the mathematics achievement levels on the national assessment of educational progress? *Applied Measurement in Education, 22*(4), 376–393.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review, 22*, 481–502.

Harmouch, T., Khraibani, Z., & Atrissi, T. (2017). A multilevel analysis to analyse times data: A comparison of the Lebanese and Singapore. *International Journal of New Technology and Research, 3*(7), 87–102.

Hastedt, D. (2016). *Mathematics achievement of immigrant students*. Cham: Springer.

IDB Analyzer V4 (version 4.0.12.0) [computer software]. (2018). Hamburg: IEA Data Processing and Research Center. Retrieved from https://www.iea.nl/data

IEA. (2017). TIMSS 2015 international database. Retrieved from the IEA website 20170610.

Ilie, S., & Lietz, P. (2010). School quality and student achievement in 21 European countries. *Issues and Methodologies in Large-Scale Assessments, 3*, 57–84.

Ismail, N. A., & Awang, H. (2008). Assessing the effects of students’ characteristics and attitudes on mathematics performance. *Problems of Education in the 21st Century, 9*, 34–41.

Johnson, E. G. (1998). *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): Research and Development Report* (A technical report No. 98–499).ETS.

Jurdak, M. (2014). Socio-economic and cultural mediators of mathematics achievement and between-school equity in mathematics education at the global level. *ZDM, 46*, 1025–1037.

Kaleli-Yilmaz, G., & Hanci, A. (2016). Examination of the 8th grade students’ TIMSS mathematics success in terms of different variables. *International Journal of Mathematical Education in Science and Technology, 47*(5), 674–695.

Kyriakides, L., & Charalambous, C. (2005). Using educational effectiveness research to design international comparative studies: Turning limitations into new perspectives. *Research Paper in Education, 20*(4), 391–412.

Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics – Theory and Methods, 46*(2), 11341–11357.

Lim, H., & Sireci, S. G. (2017). Linking TIMSS and NAEP assessments to evaluate international trends in achievement. *Education Policy Analysis Archives, 25*(11). doi:10.14507/epaa.25.2682

Lingard, B., & Grek, S. (2006). The OECD, indicators and PISA: An exploration of events and theoretical perspectives. *ESRC/ESF Research Project on Fabricating Quality in Education Working Paper, 2*(2003), 1–44.

Lipscy, P., & Kijima, R. (2016). *The politics of international testing*. Paper presented at the Assessment Power in World Politics conference, Harvard University and International Organization, 2016. Retrieved from https://web.stanford.edu/~plipscy/PoliticsofInternationalTesting2017-04-15.pdf

Martens, K., Nagel, A. K., & Windzio, M. (Eds). (2010). *Transformation of education policy*. Basingstoke: Palgrave Macmillan.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

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

National Agency for Education. (2016). Nationella prov i grundskolans årskurs 6 och 9 [National tests in elementary school grade 6 and 9] Report 447, Stockholm. Retrieved August 29, 2017, from https://www.skolverket.se/publikationer?id=3695

National Agency for Education. (2017). TIMSS, TIMSS advanced och betygen. [TIMSS, TIMSS advanced and the grades]. Stockholm. Retrieved June 15, 2017, from https://www.skolverket.se/publikationer?id=3780

Nycander, M. (2006). *Pojkars och flickors betyg* [Grades received by boys and girls]. Uppsala, Sweden: Uppsala University, Department of Education.

OCED. (2011). *Against the odds: Disadvantaged students who succeed in school*. OECD Publishing. doi:10.1787/9789264090873-en
Rolfsman, E., Wiberg, M., & Laukaityte, I. (2013). School effectiveness in the Nordic countries in relation to PISA and TIMSS. Proceedings of the IRC–2013. Retrieved from http://www.iea.nl/sites/default/files/irc/IRC-2013_Rolfsman_etal.pdf

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis (2nd ed.). London: Sage.

Sollerman and Petterson. (2016). Med fokus på matematik. Analys av samstämmighet mellan svenska styrdokument och den internationella studien TIMSS. [Focusing on mathematics. An analysis of the alignment between the Swedish steering documents and the international TIMSS study]. Stockholm: Skolverket.

Szaleniec, H., Grudniewska, M., Kondratek, B., Kulon, F., & Pokropek, A. (2013). Results of the 2002–2010 lower secondary school leaving exams on a common scale. EDUKACJA, 1, 115–134.

Tabachnik, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston, MA: Pearson Education, Inc.

Takayama, K. (2008). The politics of international league tables: PISA in Japan’s achievement crisis debate. Comparative Education, 44(4), 387–407.

Teodorovic, J. (2011). Classroom and school factors related to student achievement: What works for students? School Effectiveness and School Improvement, 22(2), 142–151.

Wiberg, M., & Rolfsman, E. (2013). School effectiveness in science in Sweden and Norway from a TIMSS perspective. Utländsk och Demokrati, 22(3), 69–84.

Wiberg, M., Rolfsman, E., & Laukaityte, I. (2013). School effectiveness in mathematics in Sweden and Norway 2003, 2007 and 2011. Proceedings of the IRC–2013. Retrieved from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC–2013_Wiberg_etal.pdf

Wikström, C. (2005). Criterion-referenced measurement for educational evaluation and selection. No. 1. Umeå, Sweden: Umeå University, Department of Educational measurement.

Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. Prospects, 39(1), 33–46.

Yariv, F. (2018). Evidence-based decision making or a tunnel vision effect? TIMSS, problem definition and policy change in Israeli mathematics education. Critical Studies in Education. doi:10.1080/17508487.2018.1448877

Zwick, R., & Green, J. G. (2007). New perspective on the correlation of scholastic assessment test scores, high school grades, and socioeconomic factors. Journal of Educational Measurement, 44(1), 1–23.

Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. Journal of Educational Measurement, 48(2), 101–121.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. American Educational Research Journal, 42(3), 439–464.