ZDM (2018) 50:643–658
https://doi.org/10.1007/s11858-018-0935-3

ORIGINAL ARTICLE

Does schooling actually perpetuate educational inequality in mathematics performance? A validity question on the measures of opportunity to learn in PISA

Kajsa Yang Hansen[1] · **Rolf Strietholt**[1,2]

Accepted: 5 April 2018 / Published online: 12 April 2018

Abstract
International assessments of mathematics have shown persistent and widely intensified socioeconomic inequalities in achievement worldwide over the last decades. Such achievement gaps may partly be due to the differences in students' personal and family characteristics. They may also be attributed to the schooling itself if school systems provide differentiated opportunity to learn (OTL) for children from privileged versus disadvantaged backgrounds. Previous research on the mechanism of the joint relationship among socioeconomic status (SES), academic achievement, and OTL produced inconclusive results. The main aim of the present study is to test whether schooling actually perpetuates social inequality in achievement, by reanalyzing PISA data. Specifically, we scrutinize the construct validity of the OTL measure in PISA that has been used in previous research. Our analyses reveal two latent dimensions of the OTL indicators in PISA, namely an unbiased OTL dimension and a self-concept dimension. The relationship between social background and mathematics achievement was only weakly mediated by OTL, when effect of students' self-concept was controlled for. Our results suggest that the previous research finding that schooling perpetuates social gaps in mathematics performance suffers from a construct validity problem.

Keywords Opportunity to learn (OTL) · Socioeconomic status · Educational inequality · Student achievement · PISA 2012 · Comparative education · Confirmatory factor analysis · Factor score

1 Introduction

International comparative studies in different domains have shown persistent and sometimes even intensified socioeconomic inequalities in achievement worldwide (e.g., Mullis et al. 2012; OECD 2014). These achievement gaps may not be due only to the differences in student individual characteristics (e.g., cognitive abilities, grit, and health), home learning environments within families (e.g., parental education, homework support, and wealth), but also due to schooling itself. If school systems provide differentiated opportunities to learn (OTL) for children from privileged versus disadvantaged backgrounds, schooling may perpetuate social gaps in learning outcomes. We define OTL simply as the degree of exposure to the educational content being tested (see also Guiton and Oakes 1995; McDonnell 1995).

In a recent study, Schmidt et al. (2015) used international PISA data to explore whether OTL is a mediator for the effect of socioeconomic status (SES) on student achievement. Based on their findings, they claimed that schooling perpetuates social inequalities in student achievement, to the extent that "roughly a third of the SES relationship to literacy is due to its association with OTL" (p. 371). However, we identified a serious methodological flaw related to their measure of OTL that may have led to a biased overestimation of the OTL effect. Basically, we argue that the questionnaire items in PISA measure not only OTL but also students' self-evaluation of their competence in mathematics. We believe that Schmidt at al. overestimated the correlation between OTL and achievement due to this construct-irrelevant source

This paper is a close collaboration, in which both authors have equally contributed substantial parts to the research idea, analyses, writing and revisions.

✉ Kajsa Yang Hansen
kajsa.yang-hansen@ped.gu.se

[1] Department of Education and Special Education, University of Gothenburg, Box 300, 405 30 Gothenburg, Sweden

[2] Institute for Education Research and School Development, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany

of variation. The main aim of the present study is to investigate this issue empirically by replicating their study with an OTL measure we adjusted for mathematics self-concept.

In the remainder of this paper, we first briefly summarize previous research on the impact of OTL on student performance. Thereafter, we discuss a methodological issue in the wording of response scales of some of the OTL measures in the PISA 2012 study, as used by Schmidt et al. (2015). Specifically, we hypothesize that the response alternative "Know it well, understand the concept" does not capture the degree of exposure to mathematics content (i.e., the definition of OTL) but mathematics self-concept. Consequently, conclusions based upon these OTL measures would be biased. In the empirical part, we adjusted the OTL measure used in the study by Schmidt at al. by means of multidimensional confirmatory factor analyses and replicated their mediation analyses. Finally, we discuss our findings from a methodological and a substantive standpoint and suggest consequences for the wording of questionnaire items.

1.1 Opportunity to learn as a research concept

The research concept OTL was introduced as part of the early international large-scale student assessment in the 1960s and 1970s, and it is an important conceptual framework for the international studies since then (Dahllöf 1971). The simple idea is that to learn something requires opportunities to learn. For this purpose, the OTL model conceptualizes curricula as functioning on three levels, as follows. (1) The level of the intended curriculum, which is what national educational policies intend students to learn and how the education system is organized to facilitate this learning (e.g., including versus segregating students with special needs). (2) The implemented curriculum, which relates to how the educational organizations (e.g., schools) implement such goals, what is actually taught in individual classrooms, who teaches it, and how it is taught. (3) Lastly, the attained curriculum, which describes what students have actually learned (e.g., as measured by scores on standardized tests) and what they think about it (e.g., their interest) as well as the emergence of educational inequality (e.g., social gaps).

There are different refinements of the OTL concept. Schmidt and McKnight (1995) propose four central research questions that organize the concept and the interrelationships among the three levels of the curriculum, namely, *What are students expected to learn? Who delivers the instruction? How is the instruction organized? What have students learned?* These research questions relate to the three levels of the curriculum and were examined at system, school, classroom, and individual levels, making the conceptual model a network of relationships among the constructs essential for opportunity to learn, not only for individual students in their classrooms and schools, but also across different educational systems. Obviously each question can be further refined, for example, there are three major channels through which the implemented curriculum is linked to the achieved curriculum. The first channel emphasizes how the characteristics of students and their peers affect instruction quality, and both in turn affect the achieved curriculum. The second channel emphasizes the effects of teacher-related factors on the achieved curriculum: teacher characteristics influence the instructional activities, which in turn affect student achievement. Here teacher characteristics, such as education, qualifications, experiences, pedagogic beliefs, and expectations, etc., affect student achievement both directly and indirectly through instructional activities. The third channel emphasizes how organizational differentiation affects teacher resources and teaching support, which in turn affect teaching activities or the implemented curriculum.

Although the OTL concept is suitable for deriving narrow and specific research questions, Schmidt et al. (2015) argue that OTL can also be interpreted in a broader sense as exposure to contents or contact coverage, for example, mathematics content. In the present paper we follow this approach and define OTL as exposure to mathematics content. Nonetheless, the previous remarks are important as they suggest that OTL (defined as content coverage) may vary and thereby perpetuate inequality in educational outcomes. Students are selected or self-selected into schools and classrooms by organizational factors such as tracking, ability grouping and school choice, and the selection is often related to students' socioeconomic and ethnic background. This brings in equity aspects of OTL that affect both instructional quality and educational outcomes. Thus, one of the foci of the current study is to examine the extent to which schooling perpetuates educational inequality in mathematics performance in different educational systems. This research focus is operationalized by investigating the interrelationships among opportunity to learn, students' family background and their mathematics achievement (see Fig. 1). One aspect of the OTL concept represents the degree to which the content in the national curriculum is actually taught, i.e., content coverage, on which emphasis was placed in our analysis, as measures of OTL.

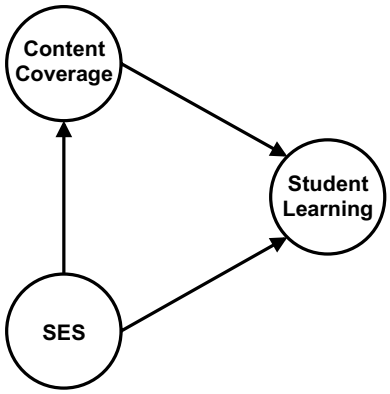

Fig. 1 The interrelationship among OTL, mathematics achievement and students' social, cultural and economic backgrounds (Schmidt et.al. 2011)

1.2 Research on the impact of opportunity to learn on student achievement

OTL has been explored in numerous studies and their findings have been incorporated in meta-analyses. In the most recent overview, Scheerens (2017) reviewed three previous meta-analyses[1] as well as 51 research studies, the majority of which were carried out in the US during the past 20 years. The three meta-analyses reported the effect sizes $d = 0.18$, 0.30, and 0.88. However, Scheerens points out that the meta-analysis that reported the largest average effect size was based on only four studies, of which two reported exceptionally large effect sizes. The overview of the 51 recent studies revealed a moderate average effect size of OTL on achievement ($d = 0.30$), which is similar to the effect sizes reported in the two other meta-analyses.

Meta-analyses are often considered to provide more generalizable findings than single studies. Hedges and Nowell, however, point out that "[r]eviews and meta-analyses of data from nonrepresentative samples are not necessarily any more representative than the studies on which they are based" (1995, p. 41). According to this, the data from many small samples of convenience should be replaced by data from large probability samples. In addition, many meta-analyses and reviews use data from the US, which limits the generalizability to other countries. For this reason, we next summarize the findings from international large-scale assessments that draw large stratified random samples from a diverse set of countries around the world.

Using PISA 2012 data from 33 OECD countries, Schmidt et al. (2015) studied the relationship between OTL, SES, and mathematics achievement, and investigated whether schooling perpetuates social inequality. They argue that the relationship between students' family socioeconomic status (SES) and their achievement is partly mediated by differences in opportunity to learn in school settings. In other words, students from privileged backgrounds attend schools that provide more learning opportunities, and these additional learning opportunities widen the existing achievement gap related to children from disadvantaged background (see Fig. 1). Schmidt at al. applied a mediation path model, where both SES and OTL have a direct effect on mathematics achievement, and SES has an additional indirect effect via OTL on mathematics achievement. The main finding of the study is a large average effect size of OTL on achievement: one unit increase in OTL corresponds with three-fifths of a standard deviation increase in performance. With respect to social inequalities, OTL mediated roughly one-third of the SES achievement gap on average. Such striking effects are particularly suspicious because they are not based on a small-scale high-quality program but on representative observational and cross-sectional international data. However, they are also in contrast to the findings from another international assessment.

Luyten (2017) examined the effect of OTL on mathematics achievement for the 22 countries that participated in both IEA (International Association for the Evaluation of Educational Achievement) TIMSS 2011 and OECD's (Organisation for Economic Co-operation and Development) PISA (2012). To measure OTL, TIMSS asked teachers if 19 main topics addressed by the TIMSS mathematics tests (e.g. simple linear equations and inequalities; congruent figures and similar triangles) were mostly taught before this year (1), mostly taught this year (2), or not yet taught or just introduced (3). Using simple regression, Luyten regressed mathematics achievement on this OTL measure. Interestingly and in contrast to the previously reported PISA findings, the average effect size of OTL on achievement was close to zero (standardized regression coefficient was 0.03). The difference between analyses based on TIMSS and PISA cannot be explained by differences in the composition of countries in both assessments because Luyten replicated the large effect sizes reported by Schmidt et al. using data only from those countries that participated in both assessments (standardized regression coefficient was 0.37; this coefficient can be transformed to Cohen's $d = 0.80$). Furthermore, Luyten aggregated the student-level OTL data from PISA on school-level to increase the comparability with TIMSS. Luyten (2017) hypothesized that the contradictory findings may be due to the fact that "TIMSS OTL measures were based on teacher responses and the PISA OTL measures on student responses" (p. 111). However, we propose an alternative explanation that is related to the response scales

[1] We disregard a fourth meta-analyses discussed by Scheerens because it operationalized OTL as enrichment programs for gifted children.

Table 1 The PISA items on content coverage in algebra and geometry

Thinking about mathematical concepts: how familiar are you with the following terms? *(Please tick only one box in each row.)*

	Never heard of it	*Heard of it once or twice*	*Heard of it a few times*	*Heard of it often*	*Know it well, understand the concept*
Exponential functions	☐1	☐2	☐3	☐4	☐5
Quadratic functions	☐1	☐2	☐3	☐4	☐5
Linear equations	☐1	☐2	☐3	☐4	☐5
Vectors	☐1	☐2	☐3	☐4	☐5
Polygons	☐1	☐2	☐3	☐4	☐5
Congruent figures	☐1	☐2	☐3	☐4	☐5
Cosines	☐1	☐2	☐3	☐4	☐5

PISA administrated computer- and paper-based surveys with marginally different layouts and instructions (see p. 170, OECD 2014)

of the PISA items. In the next section, we will scrutinize the PISA items and elaborate on this idea.

1.3 Scrutinizing how opportunity to learn is measured in PISA: a methodological issue in the wording of the response scale

To understand our critique of PISA's OTL measure, it is pertinent to revisit some foundations of the science of asking questions. Survey experts have pointed out that the question answering process involves various tasks and that apparently minor issues in the question wording or the response scale can have serious consequences for the meaning of survey data. Respondents have to understand a question, recall the information, form their judgments, translate it to the response alternatives, and edit their final answer (e.g., Schaeffer and Presser 2003; Schwarz et al. 2008). For the present study it is important to stress that the response alternatives may affect how respondents understand a question. Following Grice's (1975) maxim of relation, respondents presume that the developers of a questionnaire constructed meaningful response alternatives that are related to the task at hand.

A closer look at how PISA measures OTL reveals a methodological flaw in the wording of the instrument. Schmidt et al. (2015, p. 373) explain that they aim to measure "the intensity of exposure to selected mathematics topics." However, some of the items in the PISA questionnaire use a response scale that does not inquire only about the exposure to mathematical contents (OTL), but also about mathematical self-concept. In brief, PISA asked students how familiar they are with different mathematical terms, e.g., exponential functions (see Table 1 for the full question). The response alternatives are (1) never heard of it, (2) heard of it once or twice, (3) heard of it a few times, (4) heard of it often, and (5) know it well, understand the concept (see Table 1 for details). While the first four response categories indeed assess the intensity of exposure, the last does not.

Specifically, our critique relates to the last response alternative. While the first four response alternatives are about the frequencies of exposure to different algebra and geometry topics, the fifth one involves students' self-evaluation of their competence in these topics. Therefore, the instrument not only captures OTL but also an unintended component that we refer to as mathematical self-concept in the following. We believe that the fifth response alternative brings errors into the measurement of the latent construct OTL. Consequently, any further analyses based upon this instrument may be biased. In this vein, we hypothesize that the remarkably high association between OTL and achievement reported by Schmidt et al. (2015) is due to the biased OTL measure in PISA. Our approach to test this is to use a direct measure of mathematical self-concept to adjust the OTL measure for the unintended source of variation. Such a scale of students' self-concept in mathematics was indeed included in the PISA questionnaire and it explains on average 17.1 percent in the variance of mathematics achievement in OECD countries (OECD 2013), which corresponds to a correlation of 0.41. Based on this finding it is likely that Schmidt et al. (2015) overestimated the role of schooling in perpetuating social inequality in mathematics performance.

1.4 The present study

The main purpose of the present study is to estimate whether OTL mediates the relationship between socioeconomic background and student achievement in mathematics using PISA data for a wide range of countries. Specifically, we replicate a previous study (Schmidt et al. 2015) of this issue with an adjusted measure of OTL. For this purpose, we use multidimensional confirmatory factor analysis (CFA) and compute an unbiased OTL measure by modeling construct irrelevant variance as a

Table 2 The PISA item on formal mathematics skills problems

Below are examples of another set of mathematics problems.

1) Solve 2x + 3 = 7.
2) Find the volume of a box with sides 3m, 4m and 5m.

We want to know about your experience with these types of problems at school. Do not solve them!

(Please tick only one box in each row.)

	Frequently	*Sometimes*	*Rarely*	*Never*
How often have you encountered these types of problems in your **mathematics lessons**?	☐1	☐2	☐3	☐4

PISA administrated computer- and paper-based surveys with marginally different layouts and instructions (see p. 171, OECD 2014)

separate latent variable. We use this measure to replicate the aforementioned analyses and to achieve an unbiased estimate of the role of OTL as a mediator for the effect of SES on student achievement.

2 Data and analytical method

2.1 Samples

PISA 2012 data were used to reexamine the measurement properties of mathematics OTL and their effect on mathematics achievement. The PISA studies test the knowledge and skills of 15-year-old students in mathematics, science and reading literacy in a 3-year cycle. In each cycle, one of the core domains is tested in detail, taking up nearly two-thirds of the total testing time. In PISA 2012, mathematics was the key domain tested. It should be noted that PISA is not aligned with any specific national curriculum. Instead, it tests knowledge and abilities required in the society for the future life of the students. In the current analysis, 62 countries were included, and almost a half million students were examined.

It can be observed that countries vary greatly in their sample size. Mexico has the largest sample size, followed by Italy, Spain and Canada, with Liechtenstein's sample being the smallest. It should also be noted that PISA samples are based on age and thus they cover students from different grades. This age-sampling implies that students in the same sample had different topic coverage in mathematics because they were enrolled in different grades. It should be noted that the participation rates in PISA were below 80 percent in 1 out of 4 countries, and some countries, for example, excluded students in special education. Furthermore, the cross-cultural validity of survey data is contentiously an under-researched issue in comparative research (see Johansson 2016; Rutkowski and Rutkowski 2016; Strietholt et al. 2013; Strietholt and Scherer 2017).

2.2 Variables

2.2.1 Opportunity to learn

Three variables were used to measure OTL. The first two components were scales that summarize the mathematical content items that are related to algebra and geometry (see Table 1). The algebra scale (ALGE) is defined as the mean of three items on exponential functions, quadratic functions, and linear equations; and the geometry scale (GEOM) is defined as the mean of the next four items (vectors, polygons, congruent figures, cosines). It should be noted that the first four alternatives of the response scale are about the degree of exposure to different mathematical content while the last alternative, "Know it well, understand the concept", refers to a self-evaluation on how well the student understands the concept. The items were recoded to range from 0 to 4.

The third component of the OTL measure is a single item on how often students encounter formal problems in their mathematics lessons (FORM). Table 2 shows that two examples, an equation with one unknown and one on the volume of a box (not to be solved), define formal problems in mathematics. The response format of this single item is a frequency scale without any references to self-evaluation. The item was recoded to range from 0 to 3.

2.2.2 Mathematics self-concept

We use explicit measures of mathematics self-concept to model the construct irrelevant variance in some of the PISA OTL measures. For this purpose, we use the eight items from the student questionnaire that are listed in Table 3. To reduce the number of parameter estimates we used parceling and computed three items parcels, MSC1, MSC2, and MSC3. For the first parcel the first three items were averaged, for the second one the next three items, and for the third one the last two items. The items were recoded to range from 0 to 3 before the parceling.

Table 3 The PISA items on mathematics self-concept

How confident do you feel about having to do the following mathematics tasks?
(Please tick only one box in each row.)

	Very confident	*Confident*	*Not very confident*	*Not at all confident*
Using a train schedule to figure out how long it would take to get from one place to another.	☐1	☐2	☐3	☐4
Calculating how much cheaper a TV would be after a 30% discount	☐1	☐2	☐3	☐4
Calculating how many square feet of tile you need to cover a floor.	☐1	☐2	☐3	☐4
Understanding graphs presented in newspapers.	☐1	☐2	☐3	☐4
Solving an equation like 3x+5 = 17.	☐1	☐2	☐3	☐4
Finding the actual distance between two places on a map with a 1:10,000 scale.	☐1	☐2	☐3	☐4
Solving an equation like 2(x+3) = (x+3)(x-3).	☐1	☐2	☐3	☐4
Calculating the gas mileage of a car.	☐1	☐2	☐3	☐4

PISA administrated computer- and paper-based surveys with marginally different layouts and instructions (see p. 170, OECD 2014)

2.2.3 Socio-economic background

To measure socio-economic background, the PISA data include the index of economic, social, and cultural status (ESCS; for details see OECD 2014, pp. 351–352). This index was created on the basis of student reported information on parental occupational status, parental level of education, household wealth, and possession of educational and cultural resources.

2.2.4 Mathematics achievement

The main outcome variable is the mathematics achievement score in PISA (MATH). The achievement scale has an OECD mean of 500 with a standard deviation of 100 points. In all analyses, we used the five plausible values and combine the results using Rubin's (1987) rules.

2.2.5 Bivariate correlations with achievement

Table 4 shows that statistically significant correlations between mathematics achievement and the indicators of OTL (ALGE, GEOM, FORM), mathematic self-concept (MSC1, MSC2, MSC3), and ESCS can be observed in almost all countries. In general, mathematics achievement correlates more highly with the mathematics self-concept measures than with the respective OTL measures or the ESCS index. It should be noted that among the three OTL measures, the correlations are especially low for FORM. A possible explanation for this comparatively low correlation is that the wording of the response scale of the two other two indicators (ALGE and GEOM) led to an overestimation of the correlation with achievement. To test this possibility, we next outline our approach to remove construct irrelevant variation from the OTL measure, namely mathematics self-concept, by means of latent variable modelling.

2.2.6 Analytical method and process

Confirmatory factor analysis (CFA, see e.g., Brown 2014) was used to measure the psychometric properties and to examine construct validity of the PISA OTL instrument. A measurement model can be estimated in CFA through relating an unobserved construct, the so-called latent variable or factor, e.g., OTL, to its observed indicators, e.g., ALGE, GEOM, and FORM. As was mentioned previously, the response alternatives in ALGE and GEOM included a mathematics self-concept component. If not removed, this will bias the measurement property of the OTL construct and its effect on mathematics achievement. It also is arguable that the high OTL effect on mathematics achievement observed by Schmidt et al. (2015) might be attributed to the unintended involvement of the self-concept construct in the OTL indicators. In the current study, the measurement model of OTL was therefore specified in two ways, as an unadjusted model as replication of Schmidt et al. (2015), and as an adjusted model where the mathematics self-concept is singled out. Figure 2 illustrates the structure of the two measurement models.

The unadjusted model (2a) is a simple model with only one latent variable, OTL, indicated by ALGE, GEOM, and FORM. The adjusted model, on the other hand, accounted for the multidimensionality in the OTL measures by introducing a student's mathematics self-concept factor indicated

Table 4 Correlation of mathematics achievement with indicators of OTL, academic self-concept in mathematics, and ESCS

		ALGE	GEOM	FORM	MSC1	MSC2	MSC3	ESCS
ARE	United Arab Emirates	0.34**	0.24**	0.27**	0.32**	0.37**	0.27**	0.32**
ARG	Argentina	0.29**	0.33**	0.16**	0.24**	0.24**	0.19**	0.44**
AUS	Australia	0.50**	0.45**	0.31**	0.57**	0.57**	0.47**	0.39**
AUT	Austria	0.30*	0.37**	0.23**	0.46**	0.50**	0.38**	0.37**
BEL	Belgium	0.31**	0.41**	0.35**	0.46**	0.45**	0.32**	0.43**
BGR	Bulgaria	0.18**	0.37**	0.14**	0.23**	0.35**	0.26**	0.46**
BRA	Brazil	0.39**	0.34**	0.05**	0.24**	0.30**	0.22**	0.39**
CAN	Canada	0.31**	0.34**	0.24**	0.53**	0.53**	0.42**	0.31**
CHE	Switzerland	0.22**	0.26**	0.29**	0.56**	0.53**	0.40**	0.34**
CHL	Chile	0.36**	0.43**	0.23**	0.41**	0.42**	0.31**	0.57**
COL	Colombia	0.37**	0.29**	0.26**	0.23**	0.18**	0.14**	0.41**
CRI	Costa Rica	0.29**	0.32**	0.18**	0.18**	0.25**	0.16**	0.42**
CZE	Czech Republic	0.27**	0.42**	0.13**	0.56**	0.53**	0.38**	0.44**
DEU	Germany	0.38**	0.34**	0.25**	0.53**	0.53**	0.38**	0.42**
DNK	Denmark	0.31**	0.35**	0.22**	0.53**	0.50**	0.35**	0.42**
ESP	Spain	0.32**	0.42**	0.12**	0.45**	0.46**	0.34**	0.37**
EST	Estonia	0.10**	0.24**	0.18**	0.54**	0.46**	0.32**	0.30**
FIN	Finland	0.23**	0.32**	0.30**	0.50**	0.48**	0.37**	0.36**
FRA	France	0.25**	0.20**	0.26**	0.53**	0.52**	0.37**	0.48**
GBR	United Kingdom	0.44**	0.41**	0.30**	0.56**	0.57**	0.46**	0.36**
GRC	Greece	0.03	0.27**	0.04*	0.44**	0.48**	0.35**	0.39**
HKG	Hong Kong-China	0.28**	0.44**	0.00	0.56**	0.53**	0.47**	0.27**
HRV	Croatia	0.21**	0.22**	0.18**	0.51**	0.55**	0.39**	0.34**
HUN	Hungary	0.07**	0.39**	0.17**	0.60**	0.56**	0.42**	0.47**
IDN	Indonesia	0.08**	0.28**	0.16**	0.20**	0.20**	0.059**	0.28**
IRL	Ireland	0.38**	0.35**	0.22**	0.54**	0.52**	0.40**	0.39**
ISL	Iceland	0.02	0.24**	0.32**	0.51**	0.53**	0.42**	0.27**
ISR	Israel	0.19**	0.40**	0.06**	0.42**	0.44**	0.35**	0.41**
ITA	Italy	0.16**	0.36**	0.25**	0.48**	0.46**	0.36**	0.29**
JOR	Jordan	0.38**	0.30**	0.28**	0.18**	0.32**	0.28**	0.27**
JPN	Japan	0.18**	0.36**	0.28**	0.57**	0.57**	0.48**	0.32**
KAZ	Kazakhstan	0.12**	0.25**	0.15**	0.27**	0.22**	0.23**	0.28**
KOR	Korea	0.23**	0.51**	0.23**	0.61**	0.59**	0.55**	0.32**
LIE	Liechtenstein	0.20**	0.43**	0.26**	0.60**	0.58**	0.39**	0.30**
LTU	Lithuania	0.19**	0.37**	0.23**	0.48**	0.48**	0.36**	0.37**
LUX	Luxembourg	0.21**	0.32**	0.30**	0.50**	0.50**	0.35**	0.43**
LVA	Latvia	0.06**	0.26**	0.15**	0.47**	0.47**	0.32**	0.38**
MAC	Macao-China	0.42**	0.36**	0.00	0.52**	0.46**	0.43**	0.16**
MEX	Mexico	0.32**	0.33**	0.28**	0.35**	0.27**	0.25**	0.32**
MNE	Montenegro	0.18**	0.28**	0.15**	0.26**	0.30**	0.24**	0.34**
MYS	Malaysia	0.37**	0.35**	0.31**	0.38**	0.38**	0.26**	0.37**
NLD	Netherlands	0.49**	0.51**	0.35**	0.50**	0.41**	0.37**	0.33**
NZL	New Zealand	0.52**	0.46**	0.33**	0.55**	0.56**	0.44**	0.42**
PER	Peru	0.40**	0.36**	0.27**	0.19**	0.19**	0.19**	0.48**
POL	Poland	0.24**	0.26**	0.22**	0.62**	0.59**	0.45**	0.42**
POR	Portugal	0.33**	0.31**	0.07**	0.63**	0.59**	0.49**	0.45**
QAT	Qatar	0.43**	0.33**	0.34**	0.23**	0.31**	0.23**	0.24**
ROU	Shanghai-China	0.24**	0.27**	0.28**	0.58**	0.52**	0.55**	0.39**
RUS	Romania	0.28**	0.29**	0.18**	0.31**	0.33**	0.24**	0.43**
SGP	Russian Federation	0.06**	0.17**	0.11**	0.46**	0.43**	0.37**	0.34**
SHA	Singapore	0.59**	0.39**	0.00	0.53**	0.52**	0.49**	0.38**

Table 4 (continued)

		ALGE	GEOM	FORM	MSC1	MSC2	MSC3	ESCS
SRB	Serbia	0.12**	0.29**	0.00	0.41**	0.39**	0.30**	0.33**
SVK	Slovak Republic	0.23**	0.33**	0.20**	0.55**	0.54**	0.35**	0.50**
SVN	Slovenia	0.07**	0.19**	0.19**	0.47**	0.41**	0.32**	0.36**
SWE	Sweden	0.03	0.07**	0.22**	0.53**	0.48**	0.38**	0.32**
TAI	Chinese Taipei	0.28**	0.46**	0.16**	0.63**	0.66**	0.61**	0.42**
THA	Thailand	0.31**	0.35**	0.25**	0.26**	0.35**	0.19**	0.46**
TUN	Tunisia	0.10**	0.23**	0.21**	0.24**	0.30**	0.27**	0.35**
TUR	Turkey	0.07**	0.22**	0.08**	0.43**	0.44**	0.35**	0.39**
URY	Uruguay	0.38**	0.40**	0.00	0.35**	0.34**	0.21**	0.47**
USA	USA	0.33**	0.39**	0.22**	0.50**	0.52**	0.40**	0.39**
VNM	Viet Nam	0.18**	0.32**	0.07**	0.47**	0.43**	0.36**	0.37**

Pearson's correlation weighted by sampling weighs; * 5% significance level; ** 1% significance; level

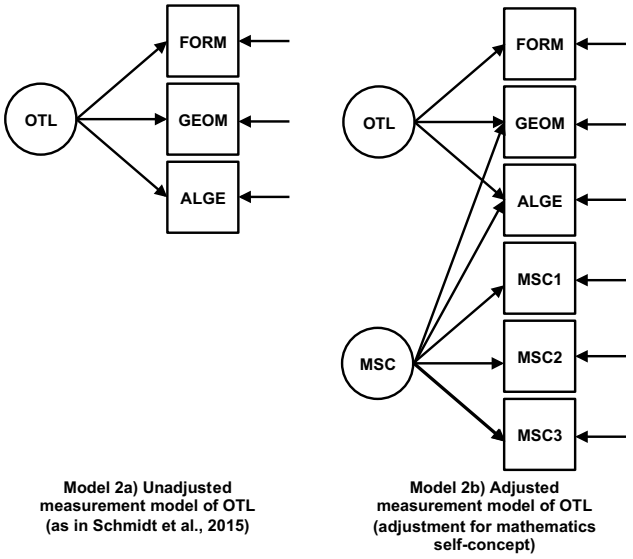

Fig. 2 Unadjusted (2a) and adjusted (2b) measurement models of OTL for the estimation of factor scores

by MSC1, MSC2, and MSC3 as well as by ALGE and GEOM (2b).

The models were evaluated by goodness of fit indices. As Hu and Bentler (1999) suggested, combinations of incremental fit indices, such as Comparative Fit indices (CFI), and absolute fit indices, such as Root Mean Square Residual of Approximate (RMSEA) and Standardized Root Mean Square Residual (SRMR), were selected to be the key model fit criteria. The cutoff values recommended by Hu and Bentler (1999) for acceptable model fit were less than 0.08 for SRMR and RMSEA. A value of less than 0.06 for RMSEA is regarded as indicating a well-fitting model, with 0.08 being acceptable. For the CFI, values greater than 0.95 are generally considered a good fit (see also Brown 2014). Since the unadjusted model is a just identified model with zero degrees of freedom, the model has a perfect model fit, in which RMSEA and SRMR are 0 and CFI is 1. In the adjusted model, an acceptable model fit has been achieved (see Table 3).

Two sets of factor scores for the OTL latent variable estimated by the adjusted and unadjusted models were saved and further used as an observed variable of OTL to examine its effect on mathematics achievement, controlling for social, cultural and economic differences in family background (see Fig. 1). It is important to note that the factor scores were computed from the pooled data of all countries, assuming measurement invariance (Schmidt et al. 2015, make the same assumption). One advantage of this approach is that the countries use the same measures of OTL, so that factor score estimates were based on the same weight (i.e., factor loadings). Another advantage of the factor score approach is simplicity. It reduces the complexity of model structure so that both direct and indirect effects can be scrutinized in a path model. The indirect effect of ESCS on mathematics achievement, mediating through OTL, can be argued as the compensatory effect of schooling on students' family background, on their academic achievement, thus indicating the degree of educational equity in different school systems.

The analytical work was done in Mplus version 7.4 (Muthén and Muthén 1998–2016), and the model complex option was applied in the modeling procedure to correct the underestimation of the standard error caused by the cluster sampling design in PISA studies (OECD 2014).

3 Results

In this section, the results from each of the analytical steps are presented. Since there are 62 countries involved in the analysis, we focus on the general patterns. Tables 6 and 7 in "Appendix" list the numerical results for the mediation analyses for each country.

Table 5 Factor loadings and model fit indices in the unadjusted model and the adjusted model with the adjustment of mathematics self-concept

	Unadjusted model		Adjusted model			
	OTL		OTL		MSC	
	Λ	t value	λ	t value	λ	t value
FORM	0.24	48.75	0.17	33.72		
ALGE	0.84	95.19	0.78	57.00	0.37	83.62
GEOM	0.69	82.48	0.57	50.14	0.36	82.53
MSC1					0.74	297.61
MSC2					0.87	443.56
MSC3					0.76	327.58
Chi-Square (df)	0 (0)		3759.086 (7)			
RMSEA[CI]	0 [0, 0]		0.033 [0.033, −0.034]			
CFI	1		0.98			
SRMR	0		0.06			

Data source is pooled international data weighted by sampling weights

3.1 Unadjusted and adjusted factor scores of opportunity to learn

Table 5 presented factor loadings estimated in both the unadjusted and the adjusted measurement models of OTL. For all three indicators of OTL, factor loadings are higher for the familiarity with different geometry and algebra concepts (i.e., GEOM and ALGE), but rather low for the experience of word problems in formal mathematics lessons or tests (FORM). When we compared the estimates between the unadjusted and the adjusted models, the factor loadings were higher in the unadjusted models. The estimates were 0.84 for ALGE, 0.69 for GEOM and 0.24 for FORM in the unadjusted model, while the average factor loadings were 0.78, 0.57 and 0.17 for ALGE, GEOM and FORM respectively in the adjusted model when the mathematics self-concept component was singled out. It should also be mentioned that rather substantial factor loadings were observed of ALGE and GEOM on the mathematics self-concept factor MSC. The decrease of the factor loadings of OTL indicators from the unadjusted model to the adjusted one indicates that the OTL indicators measured different concepts, OTL and MSC.

3.2 Opportunity to learn as a mediator of the relationship between SES and mathematics achievement

Factor scores estimated by both the adjusted and unadjusted OTL measurement models were used in a path model to replicate the analyses by Schmidt et al. (2015) and to investigate the role of OTL as a mediator for the effect of SES on student achievement, as suggested by Schmidt et al. (2011, see also Fig. 1). The parameter estimates for the direct effects are the coefficients in the regressions of MATH on ESCS, OTL on ESCS, and MATH on OTL (see Table 6 in "Appendix"). The total direct effect is the regression coefficient between ESCS and MATH, where MATH is the dependent variable affected by ESCS. The total indirect effect (i.e., the effect of ESCS on MATH mediated by OTL) can simply be calculated as a product of the two regression coefficients, namely, the regression parameter of OTL on ESCS times the parameter for the regression of MATH on OTL. The estimates of the direct and indirect effects of ESCS on mathematics achievement can be found in Table 7 in "Appendix".

The average indirect effect is higher in the model with OTL factor scores estimated by the unadjusted model, at 0.10. It reduced to 0.04 in the model where OTL factor scores were estimated while adjusting for the student's mathematics self-concept. The direct effect was found to follow an opposite change, being higher in the adjusted model (0.33) and lower in the unadjusted model (0.27). Great variation in the estimation of direct and indirect effect was also observed. In the adjusted estimates, the highest effect of ESCS on student's mathematics performance was found in the Slovak Republic, where ESCS explained 22% of the differences among students' mathematics achievement, while in Macao-China only about 2% of these differences were attributable to ESCS.

Focusing on the relationship between students' family background and PISA performance as it is mediated by OTL, i.e., the indirect effect of ESCS on mathematics achievement via OTL, Fig. 3 shows the proportion of the direct and indirect effect on mathematics achievement with OTL factor scores estimated in both the adjusted and unadjusted models. Singapore, Australia, Liechtenstein, Austria, the Netherlands and Germany, for example, have rather high indirect effects, which implies that their schools perpetuate students' family background through differentiated learning opportunities. In countries on the left-hand side of the figure, such as Iceland, Sweden,

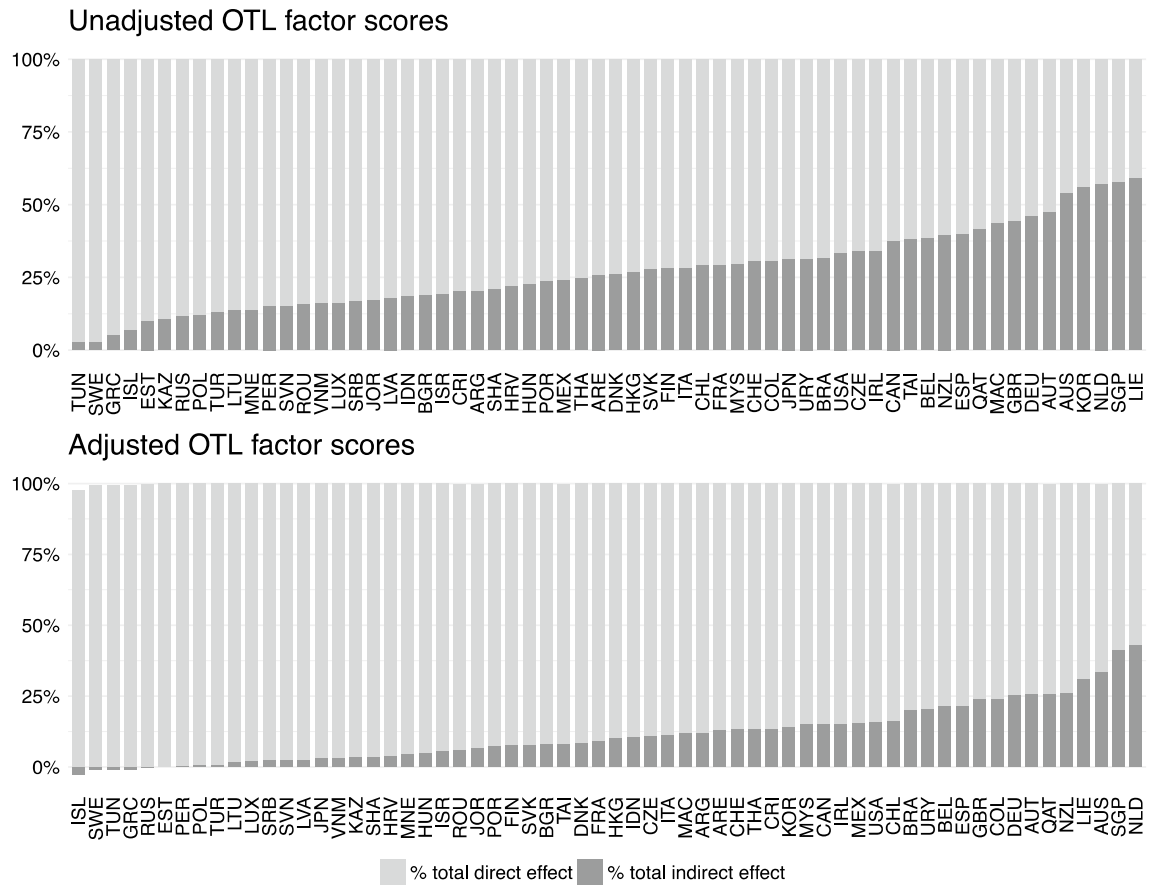

Fig. 3 Differences in the proportion of direct and indirect effects in the total effect of ESCS on mathematics achievement for models with unadjusted and adjusted OTL factor scores

Tunisia, Greece, Russia, Peru and Poland, the proportion of the mediation effect was extremely low. Some indirect effects are even negative indicating that schooling reduced social inequality. However, since the standardized effect sizes are very small (≤ 0.01) and mostly statistically non- significant, we caution against over-interpreting this finding.

The indirect effects estimated by the unadjusted model were much higher compared to those estimated by the adjusted one. It can be argued that the indirect effect achieved by the unadjusted estimation of OTL factor scores might be overestimated, since part of the effect can be attributed to the unintended component of student's mathematics self-concept. The average indirect effect of ESCS on mathematics achievement decreased by about 60%. In a small number of countries, including the Netherlands and Singapore, an increase in the indirect effect was found (see Table 7 in "Appendix"). Again, these were exceptions and we caution against over-interpreting these findings.

4 Discussion and conclusions

The current study focuses on one of the aspects in the conceptualization of OTL to examine whether or not "students have had the opportunity to study a particular topic or learn how to solve a particular type of problem presented by the test" (Husén 1967, pp. 162–163). In that definition, the content coverage or exposure to students is the key focus of the measurement of OTL. Another focus, in the current study, is to investigate OTL's mediation effect of educational equity. Applied to data from the PISA 2012 study, the two-fold aims are investigated with multidimensional confirmatory factor analysis and a factor score approach.

In the context of international large-scale studies, OTL is typically measured by teachers' perceptions of the extent to which certain topic areas within a domain were taught at school. It should be noted, however, that OTL measures in PISA 2012 were collected as students' perceptions and

experiences of some topics in different mathematics subdomains that were exposed or encountered. The information from these students deviates from the typical OTL indicators in ways that threaten the construct validity of OTL, and also the inference credibility of the impact of OTL on students' achievement and educational equity. This issue may also have profound consequences for educational practices and policy making.

4.1 Construct validity of opportunity to learn

Scrutinizing the instrument of opportunity to learn in PISA 2012, the current study found a construct-irrelevant component of students' mathematics self-concept involved in the measurement of OTL. This construct-irrelevant component was due to one of the response alternatives in the subscales of OTL, namely the exposure of different topics and concepts in algebra and geometry. We thus argued that the self-concept component should be singled out from the OTL measures so that an unbiased OTL construct can be obtained. However, we observed that this construct validity problem has not been fully addressed in previous research.

By fitting a multidimensional confirmatory factor model, the study attempted to achieve an unbiased OTL construct by removing the unintended variation of student's academic self-concept from the measures of OTL. A substantial amount of unintended variance was found in the measures of algebra and geometry subscales of OTL. The students' mathematical self-concept factor accounted for about 13% of the variation in algebra and over 14% of the variation in geometry. Measured by the residual variance of these two OTL indicators, i.e., the adjusted indicators of algebra and geometry subscales, together with the measure of students' experience with different formal mathematical concepts, the confounding component of student's academic self-concept can be controlled for. We also observed that the algebra and geometry topic exposure indicators related to the latent construct OTL highly, even after controlling for students' mathematical self-concept, although the size of the factor loadings reduced somewhat, compared to the factor loading estimated by the replication of Schmidt et al.'s (2015) conception of OTL in the unadjusted model. However, the students' experience of different mathematics concepts indicator of OTL did not relate to the latent variable OTL as highly as the other two OTL indicators.

The empirical evidence points to crucial validity problems of the OTL measure in the PISA 2012 study. The unintended component of students' academic self-concept caused by the response alternatives in OTL indicators blurred the conceptualization and measurement of OTL. Since students' academic self-concept is positively related to their achievement level (e.g., Hattie 2009; Marsh and O'Mara 2008; Möller et al. 2009), and their family background (e.g., Mruk 2006; Twenge and Campbell 2002), such a biased OTL measure will lead to an overestimation of the mediation effect. The opposite pattern of changes in the direct versus indirect effect of ESCS on mathematics achievement after adjusting for students' academic self-concept supports this interpretation.

Our analyses of the OTL measure in the PISA 2012 study illustrate how future secondary analyses of this data should be done to deal with the methodological issues related to the wording of the response scale. It is worthwhile emphasizing that it might also be useful to explore other analytical strategies to model the OTL data. The basic idea of modeling two latent factors is that students combine their perception of two continuous constructs, the degree of exposure to OTL and mathematical self-concept, when addressing items on the PISA OTL instruments. However, it may also be that students do not read all but only some response alternatives and then extrapolate the others. For example, a student who reads only the last response alternative "know it well, understand the concept" may extrapolate that the questions is only about mathematical self-concept. To gain a deeper understating of how students understand the PISA OTL instrument, think aloud protocols may provide a valuable source of information. Such qualitative data may also be used to develop and test hypotheses about student characteristics that are related to how students interpret survey questions (e.g., students with low and high self-concept may interpret questions differently).

Instead of modeling unintended sources of variance, the instruments of future PISA studies may also consider changing the wording of the response scale. For example, instead of using a five-point scale with the response alternative, "Never heard of it", "Heard of it once or twice", "Heard of it a few times", "Heard of it often", and "Know it well, understand the concept", the developers of the instruments for future PISA cycles may simply remove the last alternative for a clearer interpretation of the items.

Another validity problem related to the OTL indicator students' experiences of mathematics concepts is that such experiences can occur in different intertwined contexts in everyday life at home, with peers, in leisure time activities and at school. It is hard to separate the mathematics related experiences that are particularly connected to school settings from those encountered elsewhere. The fact that children's experiences of knowledge and learning are constrained by their family socioeconomic background (e.g., Evans 2004) may therefore introduce an additional unintended variation into OTL measures. This may also explain why OTL was substantially related to student's achievement in PISA but not in TIMSS (Luyten 2017). Another approach to achieve valid OTL measures may be to survey teachers who have the direct perception of their practices, as in, for example, IEA TIMSS studies.

4.2 Does schooling actually perpetuate educational inequality in mathematics performance?

Factor scores estimated by both the unadjusted and the adjusted model of OTL were further used to replicate Schmidt et al.'s (2015) analyses and to investigate the role of OTL as a mediator for the effect of SES on student mathematics achievement. The effects of OTL were shown to be higher when the unadjusted model was used to compute OTL scores, compared to those estimated by the adjusted model. After parsing the mathematics self-concept component from the OTL measures, and controlling for the family background differences, we observed a smaller OTL effect on mathematics achievement as well as a smaller indirect effect of family background on performance.

There has been a long-standing discussion of whether school makes a difference in students' academic performance when taking into account family background (e.g., Coleman et al. 1966; Gamoran and Long 2007). While empirical evidence has consistently shown that students' SES affects their academic achievement (e.g., Sirin 2005), the effects of school-related factors are at best weak, and often not significant. After adjusting for students' mathematics self-concept, the current study confirmed that students' socioeconomic status is positively related to their mathematics achievement in all the countries studied, with the same being true for OTL. We also found an additional positive and significant indirect effect of family background on mathematics achievement mediating through the differences in opportunity to learn. It seemed that schools do not close, but they widen, the social achievement gap. However, compared to the effects achieved in the Schmidt et al. study (2015), the effect size of the indirect effects is much lower, indicating an overestimation in the previous study by Schmidt and colleagues. It should be noted that the great variation observed in the indirect effect of students' family social, cultural and economic background on their mathematics achievement mediating students' educational opportunity indicates that different school systems maintain the socioeconomic status of their children to a varying degree. Even though the majority of the countries we studied showed a significant indirect effect, the effects are rather small in most of the countries. Moreover, due to the validity problems in the OTL in the PISA study, as we discussed before, the interpretation of the indirect effect needs to be cautious. It can be hypothesized that schooling may be differentiated for students from different social, economic and cultural background (see, e.g., Yang Hansen and Gustafsson 2016), schools may not only mediate the student's family background but also compensate or anti-compensate the relationship between student's family background and their academic achievement (Gustafsson et al. 2016). In order to offer a relatively complete picture of the question of whether schooling perpetuates educational inequality, a cross-level interaction between OTL and educational equity has to be examined.

4.3 Implications for educational practices and educational policy-making

One of the most important policy issues within the field of education is the degree to which nationally determined curricula affect the outcomes of education, and it is a challenging scientific task to determine the mechanisms involved. OTL is crucial in this context. As McDonnell (1995) pointed out, OTL is a generated research concept and policy instrument that "has changed how researchers, educators, and policy makers think about the determinants of student learning" (p. 305).

In international studies, OTL can be used not only for explaining the variation in students' test results, but also for evaluating equity and quality of the educational environments. Especially when global educational reforms emphasize accountability, the concept OTL is used as an instrument to understand between country variations in academic achievement and to guide their educational reforms and policy-making related to national curricula. Given the importance of OTL, it is essential for researches to provide credible research evidence that is based on valid conceptualization and measurement of OTL.

It is necessary to emphasize that the frameworks of the IEA and the OECD assessments are based on different principles, the IEA taking a starting point in curricular goals, and PISA focusing on competencies (McGaw 2008a, b; Wagemaker 2008). This emphasis implies that the OTL information in the PISA studies is limited. Moreover, OTL measures in PISA were collected from individual students' experiences in formative mathematics education of different mathematics concepts. One could argue that such measures may suffer from a lack of construct validity. Nevertheless, it is of interest to analyze the relations between curricula and outcomes, and to compare the IEA and OECD approaches to constructing assessment frameworks. One approach is to combine information from different studies, since many of the countries that participate in PISA also participate in TIMSS and PIRLS.

Appendix

See Tables 6 and 7.

Table 6 Direct effects between ESCS, OTL and mathematics performance

Country	Estimated of the model with unadjusted OTL factor scores						Estimated of the model with adjusted OTL factor scores					
	Math-OTL		Math-ESCS		OTL-ESCS		Math-OTL		Math-ESCS		OTL-ESCS	
	β	t	β	t	β	t	β	t	β	t	β	t
ARE	0.42	31.90	0.23	14.36	0.20	13.78	0.31	21.13	0.27	16.29	0.13	9.39
ARG	0.29	18.13	0.31	15.19	0.28	16.53	0.20	12.87	0.34	16.50	0.23	14.03
AUS	0.58	78.00	0.16	18.38	0.33	32.36	0.44	51.80	0.23	24.17	0.27	26.92
AUT	0.46	28.60	0.21	11.42	0.42	22.62	0.29	16.56	0.30	15.31	0.36	19.91
BEL	0.46	37.79	0.27	20.51	0.38	28.35	0.31	25.88	0.35	25.16	0.31	24.06
BGR	0.29	19.11	0.38	17.06	0.31	17.67	0.18	12.39	0.43	18.36	0.22	13.20
BRA	0.38	30.26	0.28	17.08	0.34	21.94	0.28	23.91	0.32	18.08	0.28	19.61
CAN	0.43	52.88	0.20	19.05	0.27	24.38	0.25	27.37	0.26	23.50	0.19	17.59
CHE	0.38	27.07	0.25	17.90	0.28	17.59	0.22	14.73	0.31	20.57	0.22	13.83
CHL	0.39	28.94	0.34	18.87	0.36	21.84	0.27	19.83	0.40	21.60	0.29	16.98
COL	0.40	28.49	0.27	15.01	0.31	15.58	0.35	26.05	0.30	15.98	0.28	14.25
CRI	0.31	17.66	0.35	15.39	0.29	12.69	0.25	14.99	0.38	16.14	0.24	10.84
CZE	0.43	28.07	0.27	14.70	0.31	17.67	0.23	14.30	0.36	17.94	0.19	10.58
DEU	0.50	36.97	0.22	15.60	0.37	25.24	0.34	23.00	0.30	20.15	0.30	20.43
DNK	0.34	23.61	0.31	20.40	0.31	20.55	0.16	12.15	0.38	25.74	0.22	14.35
ESP	0.44	47.48	0.24	23.42	0.36	37.05	0.29	30.24	0.31	29.08	0.29	28.17
EST	0.22	14.26	0.27	13.91	0.15	9.55	−0.02	−1.25	0.30	15.71	0.02	1.47
FIN	0.40	31.14	0.23	17.46	0.22	16.16	0.21	15.75	0.29	21.08	0.11	8.47
FRA	0.40	29.87	0.34	22.57	0.35	20.22	0.19	12.59	0.43	27.61	0.23	13.66
GBR	0.52	41.79	0.20	14.97	0.31	20.57	0.35	26.34	0.27	19.07	0.24	15.30
GRC	0.12	7.63	0.37	18.58	0.19	12.16	−0.07	−4.54	0.40	20.14	0.04	2.84
HKG	0.40	29.79	0.19	8.58	0.18	10.92	0.24	17.92	0.24	10.16	0.12	7.25
HRV	0.37	28.63	0.28	15.28	0.20	12.57	0.16	11.88	0.34	16.04	0.09	5.89
HUN	0.36	21.56	0.37	19.01	0.31	18.28	0.14	7.72	0.46	21.93	0.17	10.04
IDN	0.23	13.10	0.26	6.76	0.25	9.41	0.17	9.94	0.28	7.03	0.20	7.91
IRL	0.43	34.63	0.25	15.43	0.31	19.13	0.26	19.02	0.32	19.71	0.23	13.50
ISL	0.10	5.01	0.26	15.20	0.16	8.98	−0.12	−6.34	0.28	16.65	0.06	3.60
ISR	0.34	20.17	0.33	19.96	0.24	13.31	0.19	11.66	0.39	22.29	0.13	7.45
ITA	0.37	43.03	0.23	21.83	0.23	25.63	0.22	25.27	0.28	25.03	0.16	18.34
JOR	0.33	19.81	0.24	8.79	0.16	8.55	0.25	14.12	0.27	9.75	0.08	4.26
JPN	0.42	28.99	0.22	12.81	0.24	13.88	0.11	6.96	0.31	14.96	0.09	6.20
KAZ	0.17	11.36	0.25	10.67	0.20	11.95	0.08	5.35	0.28	11.39	0.12	7.24
KOR	0.53	33.45	0.14	9.58	0.34	18.97	0.24	13.28	0.28	14.74	0.19	11.89
LIE	0.46	7.84	0.11	1.85	0.35	4.21	0.30	4.90	0.18	2.85	0.28	3.04
LTU	0.29	22.82	0.31	19.61	0.18	10.77	0.09	7.07	0.36	21.44	0.08	4.78
LUX	0.21	10.21	0.36	13.54	0.33	14.24	0.04	1.80	0.42	15.42	0.22	8.67
LVA	0.29	17.48	0.32	17.83	0.23	12.50	0.10	5.45	0.38	19.90	0.10	5.66
MAC	0.53	30.47	0.09	3.57	0.12	4.38	0.34	19.92	0.14	4.50	0.06	2.12
MEX	0.36	43.71	0.25	22.64	0.23	22.96	0.27	32.31	0.27	24.33	0.19	20.92
MNE	0.28	14.73	0.31	6.47	0.17	10.07	0.16	8.42	0.34	6.88	0.10	6.20
MYS	0.42	31.67	0.26	13.77	0.26	15.19	0.30	20.77	0.31	15.52	0.18	11.44
NLD	0.62	52.37	0.15	9.59	0.32	16.22	0.50	37.88	0.20	11.38	0.30	15.31
NZL	0.52	35.11	0.26	17.88	0.33	19.78	0.41	28.57	0.32	22.24	0.27	16.10
PER	0.27	10.35	0.28	10.58	0.19	6.48	0.03	0.87	0.33	10.89	0.06	2.14
POL	0.27	17.67	0.36	20.09	0.20	11.25	0.05	2.90	0.40	21.63	0.06	3.57
POR	0.37	27.27	0.35	21.60	0.29	16.01	0.20	14.00	0.42	24.25	0.18	10.02
QAT	0.46	28.33	0.14	6.70	0.21	11.84	0.37	23.62	0.17	7.54	0.17	10.83

Table 6 (continued)

Country	Estimated of the model with unadjusted OTL factor scores						Estimated of the model with adjusted OTL factor scores					
	Math-OTL		Math-ESCS		OTL-ESCS		Math-OTL		Math-ESCS		OTL-ESCS	
	β	t	β	t	β	t	β	t	β	t	β	t
ROU	0.26	16.63	0.37	16.09	0.27	13.57	0.15	9.79	0.41	16.83	0.18	10.06
RUS	0.20	13.78	0.30	14.20	0.22	14.35	0.03	−1.70	0.35	16.17	0.03	2.05
SGP	0.60	59.15	0.16	10.79	0.37	23.43	0.50	43.22	0.22	13.94	0.32	21.22
SHA	0.31	16.86	0.30	14.86	0.27	11.85	0.09	5.35	0.37	16.14	0.15	7.51
SRB	0.26	18.99	0.29	12.86	0.21	14.17	0.07	5.19	0.34	14.40	0.11	8.15
SVK	0.37	24.12	0.36	19.75	0.37	17.44	0.16	9.53	0.46	23.19	0.24	11.24
SVN	0.24	15.16	0.33	16.84	0.24	13.37	0.07	4.35	0.38	18.46	0.12	6.67
SWE	0.06	2.68	0.32	19.99	0.14	8.55	−0.12	−5.47	0.33	20.94	0.02	1.39
TAI	0.45	34.07	0.26	17.51	0.36	24.12	0.17	10.69	0.39	23.34	0.21	13.49
THA	0.29	20.56	0.24	7.46	0.27	16.57	0.20	13.88	0.27	8.18	0.21	14.11
TUN	0.05	2.45	0.34	11.40	0.15	7.53	−0.06	−3.08	0.36	11.37	0.04	2.25
TUR	0.23	17.09	0.33	13.97	0.21	11.85	0.02	1.75	0.38	14.89	0.11	6.76
URY	0.40	28.07	0.33	17.57	0.36	19.18	0.31	21.44	0.38	19.14	0.32	16.49
USA	0.45	30.98	0.26	17.34	0.28	13.74	0.30	19.45	0.33	21.50	0.20	9.99
VNM	0.29	19.08	0.31	11.59	0.22	10.60	0.11	6.54	0.37	12.76	0.10	5.23
Mean	0.35	26.11	0.27	15.06	0.27	16.00	0.19	14.23	0.33	17.38	0.18	10.92
SD	0.12	13.87	0.07	4.80	0.07	6.40	0.14	11.64	0.07	5.74	0.09	6.46

Each country samples is weighted by sampling weighs

Table 7 Total direct and total indirect effect of ESCS on math achievement

	Estimated of the model with unadjusted OTL factor scores				Estimated of the model with adjusted OTL factor scores				Overestimation of indirect effects
	Indirect effect		Direct effect		Indirect effect		Direct effect		
	Est.	t value	Est.	t value	Est.	t value	Est.	t value	
ARE	0.08	11.93	0.23	14.36	0.04	8.31	0.27	16.29	43.21
ARG	0.08	11.75	0.31	15.19	0.05	9.22	0.34	16.50	34.02
AUS	0.19	30.05	0.16	18.38	0.12	24.55	0.23	24.17	5.75
AUT	0.19	18.03	0.21	11.42	0.10	12.67	0.30	15.31	28.43
BEL	0.17	21.28	0.27	20.51	0.10	16.51	0.35	25.16	29.73
BGR	0.09	13.32	0.38	17.06	0.04	9.48	0.43	18.36	51.69
BRA	0.13	15.59	0.28	17.08	0.08	13.81	0.32	18.08	21.71
CAN	0.12	22.10	0.20	19.05	0.05	14.85	0.26	23.50	51.83
CHE	0.11	13.14	0.25	17.90	0.05	8.90	0.31	20.57	47.85
CHL	0.14	17.40	0.34	18.87	0.08	12.93	0.40	21.60	33.70
COL	0.12	12.37	0.27	15.01	0.10	11.57	0.30	15.98	−1.87
CRI	0.09	8.87	0.35	15.39	0.06	7.76	0.38	16.14	23.33
CZE	0.14	14.52	0.27	14.70	0.04	8.21	0.36	17.94	63.41
DEU	0.19	19.14	0.22	15.60	0.10	14.29	0.30	20.15	26.53
DNK	0.11	14.11	0.31	20.40	0.04	8.87	0.38	25.74	63.91
ESP	0.16	30.04	0.24	23.42	0.09	20.32	0.31	29.08	30.57
EST	0.03	8.06	0.27	13.91	**0.00**	**−0.91**	0.30	15.71	–
FIN	0.09	14.57	0.23	17.46	0.02	7.71	0.29	21.08	69.78
FRA	0.14	17.06	0.34	22.57	0.04	9.53	0.43	27.61	66.41
GBR	0.16	17.04	0.20	14.97	0.09	12.18	0.27	19.07	30.21
GRC	0.02	6.28	0.37	18.58	0.00	−2.40	0.40	20.14	113.53
HKG	0.07	9.62	0.19	8.58	0.03	6.50	0.24	10.16	58.68

Table 7 (continued)

	Estimated of the model with unadjusted OTL factor scores				Estimated of the model with adjusted OTL factor scores				Overestimation of indirect effects
	Indirect effect		Direct effect		Indirect effect		Direct effect		
	Est.	t value	Est.	t value	Est.	t value	Est.	t value	
HRV	0.08	10.89	0.28	15.28	0.01	5.14	0.34	16.04	80.81
HUN	0.11	13.66	0.37	19.01	0.02	5.93	0.46	21.93	78.76
IDN	0.06	6.16	0.26	6.76	0.03	5.12	0.28	7.03	33.92
IRL	0.13	15.80	0.25	15.43	0.06	10.52	0.32	19.71	48.51
ISL	0.02	4.57	0.26	15.20	−0.01	−3.05	0.28	16.65	142.65
ISR	0.08	10.69	0.33	19.96	0.02	6.03	0.39	22.29	69.16
ITA	0.09	20.57	0.23	21.83	0.04	13.98	0.28	25.03	53.36
JOR	0.05	7.34	0.24	8.79	0.02	3.84	0.27	9.75	59.50
JPN	0.10	11.30	0.22	12.81	0.01	4.31	0.31	14.96	89.91
KAZ	0.03	7.92	0.25	10.67	0.01	4.37	0.28	11.39	68.72
KOR	0.18	15.01	0.14	9.58	0.05	8.24	0.28	14.74	70.76
LIE	0.16	2.91	0.11	1.85	0.08	2.00	0.18	2.85	25.38
LTU	0.05	9.40	0.31	19.61	0.01	3.85	0.36	21.44	86.57
LUX	0.07	6.69	0.36	13.54	**0.01**	**1.55**	0.42	15.42	–
LVA	0.07	9.94	0.32	17.83	0.01	4.06	0.38	19.90	84.45
MAC	0.07	4.15	0.09	3.57	0.02	2.03	0.14	4.50	66.77
MEX	0.08	20.05	0.25	22.64	0.05	17.14	0.27	24.33	24.38
MNE	0.05	10.25	0.31	6.47	0.02	5.36	0.34	6.88	63.57
MYS	0.11	12.83	0.26	13.77	0.06	9.42	0.31	15.52	40.10
NLD	0.20	14.67	0.15	9.59	0.15	13.88	0.20	11.38	−29.91
NZL	0.17	16.49	0.26	17.88	0.11	13.68	0.32	22.24	12.02
PER	0.05	5.65	0.28	10.58	**0.00**	**0.77**	0.33	10.89	–
POL	0.05	9.19	0.36	20.09	0.00	2.15	0.40	21.63	94.19
POR	0.11	14.41	0.35	21.60	0.03	8.29	0.42	24.25	64.91
QAT	0.10	10.09	0.14	6.70	0.06	9.32	0.17	7.54	15.58
ROU	0.07	10.41	0.37	16.09	0.03	7.51	0.41	16.83	58.81
RUS	0.04	10.91	0.30	14.20	**0.00**	**−1.31**	0.35	16.17	–
SGP	0.22	22.83	0.16	10.79	0.16	20.64	0.22	13.94	−19.75
SHA	0.08	10.06	0.30	14.86	0.01	4.05	0.37	16.14	82.70
SRB	0.06	10.56	0.29	12.86	0.01	4.08	0.34	14.40	85.11
SVK	0.14	13.81	0.36	19.75	0.04	6.77	0.46	23.19	69.11
SVN	0.06	9.55	0.33	16.84	0.01	3.54	0.38	18.46	84.07
SWE	0.01	2.44	0.32	19.99	**0.00**	**−1.41**	0.33	20.94	–
TAI	0.16	18.36	0.26	17.51	0.04	7.90	0.39	23.34	76.15
THA	0.08	12.51	0.24	7.46	0.04	10.39	0.27	8.18	37.00
TUN	0.01	2.08	0.34	11.40	0.00	−2.34	0.36	11.37	142.37
TUR	0.05	9.47	0.33	13.97	**0.00**	**1.72**	0.38	14.89	–
URY	0.15	15.77	0.33	17.57	0.10	13.00	0.38	19.14	15.44
USA	0.13	11.91	0.26	17.34	0.06	8.27	0.33	21.50	43.25
VNM	0.06	8.68	0.31	11.59	0.01	3.73	0.37	12.76	80.69
Mean	0.10	12.68	0.27	15.06	0.04	7.73	0.33	17.39	58.21
Min	0.01	2.08	0.09	1.85	−0.01	−3.05	0.14	2.85	−29.91
Max	0.22	30.05	0.38	23.42	0.16	24.55	0.46	29.08	142.65

Each country samples is weighted by sampling weighs; estimates in bold are not statistically significant, and the overestimation of indirect effect in the unadjusted model thus is not computed for these countries

References

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.

Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, A. M., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Pringting Office.

Dahllöf, U. (1971). Relevance and fitness analysis in comparative education. *Scandinavian Journal of Educational Research, 15*(3), 101–121.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*(2), 77–92. https://doi.org/10.1037/0003-066X.59.2.77.

Gamoran, A., & Long, D. A. (2007). Equality of educational opportunity: a 40 year retrospective. In R. Teese, S. Lamb & M. Duru-Bellat, Helme S. (Eds.), *International studies in educational inequality, theory and policy* (pp. 23–46). Dordrecht: Springer.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). New York: Academic Press.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis, 17*(3), 323–336.

Gustafsson, J.-E., Nielsen, T., & Yang Hansen, K. (2016). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2016.09.004 **(Advance online publication)**.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Husén, T. (Ed.). (1967). *International study of achievement in mathematics: A comparison of twelve countries (Vol. I)*. New York: Wiley.

Johansson, S. (2016). International large-scale assessments: what uses, what consequences? *Educational Research, 58*(2), 139–148. https://doi.org/10.1080/00131881.2016.1165559.

Luyten, H. (2017). Predictive power of OTL measures in TIMSS and PISA. In J. Scheerens (Ed.) *Opportunity to learn, curriculum alignment and test preparation* (pp. 103–119). New York: Springer International.

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*(4), 542–552.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305–322.

McGaw, B. (2008a). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy & Practice, 15*(3), 223–243.

McGaw, B. (2008b). Further reflections. *Assessment in Education: Principles. Policy & Practice, 15*(3), 279–282.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129–1167.

Mruk, C. J. (2006). *Self-esteem research, theory, and practice: Toward a positive psychology of self-esteem*. New York: Springer.

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.

Muthén, L. K., & Muthén, B. O. (1998–2016). *Mplus User's Guide* (7th edn.). Los Angeles, CA: Muthén & Muthén.

OECD. (2012). Student questionnaire Form B in PISA 2012. Document retrieved June 2016 from https://nces.ed.gov/surveys/pisa/pdf/MS12_StQ_FormB_ENG_USA_final.pdf.

OECD. (2013). *PISA 2012 results: Ready to learn: Students' engagement, drive and self-beliefs* (Volume III). Paris: OECD Publishing.

OECD. (2014). *PISA 2012 results: What students know and can do–Student performance in mathematics, reading and science (volume I, Revised edition, February 2014)*. Paris: OECD Publishing. https://doi.org/10.1787/9789264201118-en.

OECD. (2014). *PISA 2012 technical report*. Paris: OECD Publishing. Retrieved June 2016 from https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken: Wiley.

Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher, 45*(4), 252–257.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65–88.

Scheerens, J. (2017). *Opportunity to learn, curriculum alignment and test preparation: A research review*. New York: Springer International.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher, 44*(7), 371–386.

Schmidt, W. H., Cogan, L. S., Houang, R. T., & McKnight, C. C. (2011). Content coverage differences across districts/states: A persisting challenge for US education policy. *American Journal of Education, 117*(3), 399–427.

Schmidt, W. H., & McKnight, C. C. (1995). Surveying educational opportunity in mathematics and science: An international perspective. *Educational evaluation and policy analysis, 17*(3), 337–353.

Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. D. de Leeuv, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology*. New York: Lawrence Erlbaum Associates.

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

Strietholt, R., Rosén, M., & Bos, W. (2013). A correction model for differences in the sample compositions: The degree of comparability as a function of age and schooling? *Large-Scale Assessments in Education, 1*(1), 1–20.

Strietholt, R., & Scherer, R. (2017). The contribution of international large-scale assessments to educational research: Combining individual and institutional data sources. *Scandinavian Journal of Educational Research*, 1–18.

Twenge, J. M., & Campbell, W. K. (2002). Self-esteem and socioeconomic status: A meta-analytic review. *Personality and social psychology review, 6*(1), 59–71.

Wagemaker, H. (2008). Choices and trade-offs: Reply to McGaw. *Assessment in Education: Principles, Policy & Practice, 15*(3), 267–278.

Yang Hansen, K., & Gustafsson, J. E. (2016). Causes of educational segregation in Sweden–School choice or residential segregation. *Educational Research and Evaluation, 22*(1–2), 23–44.