Lafontaine, Dominique; Baye, Ariane; Vieluf, Svenja; Monseur, Christian

# Equity in opportunity-to-learn and achievement in reading. A secondary analysis of PISA 2009 data

# Equity in opportunity-to-learn and achievement in reading: A secondary analysis of PISA 2009 data

Dominique Lafontaine[a], A. Baye[a], S. Vieluf[b], C. Monseur[a]

[a] Department of Education, University of Liège, Boulevard du Rectorat, 5, B32, 4000 Liège, Belgium
[b] German Institute for International Education Research (DIPF), Schloßstraße, 29, 60486 Frankfurt am Main, Germany

## Abstract

Using data from PISA 2009, the present study investigates firstly how equally students are exposed to opportunities to improve their reading skills (OTL) depending on the school they are enrolled in, and secondly the links between OTL in reading and achievement at the school level. A multidimensional within-item IRT is used to model the OTL. The intraclass correlation of both OTL dimensions issued from the IRT analysis – reading fiction and reading non-continuous tasks – is high, especially in differentiated education systems, showing an unequal exposure to OTL in reading according to the school. Robust correlations between the two OTL dimensions and reading achievement are observed at the school level. In addition, the results of a multilevel regression analysis show that a substantial proportion of the between-school variance in reading can be explained by OTL and by the school social intake. The proportion of between-school variance explained jointly by OTL and social intake is higher in differentiated education systems than in comprehensive ones.

Keywords: Student evaluation, opportunity-to-learn, large-scale assessments, reading literacy, equity, secondary education

## Introduction

To what extent are 15-year-old students given equal opportunities to enhance their reading literacy, regardless of the school they are enrolled in? Do some education systems provide their students in secondary education with more equal opportunities to learn in reading?

The concept of opportunity-to-learn (OTL) refers to a fundamental instructional process: ''what students learn in school is related to what is taught in school'' (Schmidt & Maier, 2009, p. 541). Among the malleable variables that policy-makers may want to use to improve their education policies, opportunity-to-learn is known to be one that shows a constant positive association with achievement (Muthén et al., 1995; Schmidt & Maier, 2009; Wang, 1998). A number of international comparative studies have extensively assessed the association of OTL variables with student scores, mainly in mathematics and science. Fewer studies have explored OTL in domains such as reading and writing, and most of them were about primary education (Grisay, 2008; Leinhardt, Zigmond, & Cooley, 1981; Lundberg, 1994; Martinez, 2012). This research gap exists, even though cross-country and within-country variation exists in most instructional practices implemented in schools to teach the students the language of instruction. For instance, variation may exist in the amount of time that students spend reading literature or other kinds of texts – whether retrieving names, facts or simple information from text or developing more cognitively challenging skills such as relating parts of the text to one another, evaluating the form and content of the text or taking a critical stance. Within countries, the amount and nature of the reading tasks assigned to students may vary from school to school and from class to class, due to possible differences in local curricula and/or study programs.

This is especially true in differentiated education systems, where teachers are likely to teach more challenging skills such as the analysis of poetry and literary texts to students enrolled in academic programs, while less demanding skills – e.g. locating information in functional, work-related texts – tend to be emphasized for students enrolled in vocational tracks.

To address the issue of the equitable exposure to OTL in reading in a broad set of countries, data from the PISA 2009 study were used for secondary analyses. PISA – the Program for International Student Assessment – is a large international comparative study aimed at assessing the reading, mathematical and scientific literacy of 15-year-old students in OECD and partner countries every 3 years (for more details, see http://www.pisa.OECD.org).

## State of the art

### The OTL concept

The concept of opportunity-to-learn was first introduced by J. B. Carroll in the early 1960s. Carroll viewed OTL mainly as a measure of content exposure (allocated learning time), but OTL also refers to content coverage or content emphasis – which topics are selected for more emphasis, e.g. higher- or lower-order skills (Stevens, 1993; Wang, 1998). The development of OTL measures is closely linked with the first IEA comparative studies, assessing mathematics (FIMS, 1965) and science (FISS, 1971) (Floden, 2002; Schmidt & Maier, 2009; Schmidt & McKnight, 1995). Results from these studies provide evidence for variation in students' exposure to the curriculum, which can account for differences in achievement between and within countries.

### How OTL in reading is measured in cross-cultural contexts

As reading is not taught as a specific subject, especially in secondary education, measuring OTL is not as straightforward as it is for mathematics, for instance. One approach is to try to capture how similar the reading material and the kind of questions in the reading test (e.g. PISA or PIRLS) are to the reading assessments usually taken by students in their courses; this approach is close to the notion of content coverage. Another approach aims at identifying to what extent students have had the opportunities to learn the kind of reading strategies or processes necessary to answer the questions in the test; this approach is closer to the notions of content exposure and content emphasis.

In IEA-RLS (1991) and PIRLS (2001, 2006, 2011), teachers were asked about various aspects of OTL and instruction in reading, including reading material, teaching strategies and types of responses to the text (oral, written, quiz). Likert scales were used to measure frequency of use, exposure or emphasis. It is nowadays well established that Likert scales are especially sensitive to response style biases (e.g. Harzing et al., 2009; Kyllonen & Bertling, 2014; Rocereto, Puzakova, Anderson, & Kwak, 2011; Yang, Harkness, Chin, & Villar, 2010). Response styles can be defined as a systematic tendency to respond to items on a basis other than what the items were designed to measure; examples are acquiescence, disacquiescence, social desirability, extremity scoring or midpoint responding. Differences in response styles are not only observed between individuals, but also between cultures. For example a stronger tendency to use the intermediate levels of response scales in East Asian as compared to Western countries, in which it is more common to use the extreme points of Likert scales, has been reported by several studies (Johnson, Kulesa, Cho, & Shavitt, 2005; for a critical overview, see Yang et al., 2010). These systematic differences in response styles can lead to

misinterpretations of comparisons of cultures based on Likert scales, because observed differences may reflect cultural response tendencies rather than substantial differences. They may also be one explanation for the attitude-achievement paradox, that is, negative country-level correlations of attitudinal constructs with student achievement, in spite of positive within-country correlations (Buckley, 2009; Van de Gaer, Grisay, Schultz, & Gebhardt, 2012).

<u>OTL and equity</u>

When examining the equality of education systems, Sen's question immediately comes to mind: ''Equality of what?'' (Sen, 1992). Equality is not a unique concept, since several principles of equality coexist (equality of opportunities, equality of treatment, equality of achievement or equality of social outputs) and compete with one another (EGREES, 2005; Grisay, 1984). For instance, if equality of achievement is a goal, equality of treatment may be challenged and overturned, since strict equality of treatment may reinforce social inequalities (Dubet, 2010; Duru-Bellat, 2009; Felouzis, 2014). The concept of equity is necessary for making choices between competing principles of equality. Unlike the principle of equality, the concept of equity refers to a normative principle, a principle of justice according to which some types of inequalities are judged as fair, and others not (EGREES, 2005). Equity is a matter of justice and different theories of justice may be used to judge the fairness of education systems. For instance, Rawls' (1971) theory stipulates that, under control of certain liberties, education should foster a ''fair equality of social opportunities''. Other inequalities in education, in particular inequalities in knowledge and skills between more and less privileged students, should benefit to the latter (Meuret, 1999).

Regarding OTL, strict equality of treatment may be seen as fair: from a meritocratic point of view, each student may then take advantage of the opportunities offered, and is responsible for the development of his/her own skills, provided that the same curriculum is offered. Another way to judge the equity of distribution of OTL is to take the view that some students must be provided with more of the same OTL (according to a compensatory principle) or with different OTL (adaptivity: a curriculum tailored to individual/local needs in order to reach the same achievement standards).

Despite different views, most of the theoreticians of justice, from supporters of meritocracy to egalitarianism, would agree on one point: it hurts the idea of educational justice if OTL distribution depends on the students' socioeconomic and cultural status and if less beneficial OTL is offered to less privileged students, while more privileged students are exposed to more challenging content or goals.

For the present study, we will consider as more equitable these education systems with smaller variance between schools in OTL and where schools' social intake is not linked or is even negatively linked to OTL; the latter case (negative correlation) would reflect a policy of compensating for poor or underprivileged socioeconomic background by providing more or better OTL to disadvantaged pupils.

Since the beginning, OTL measures have been related to equity issues in the IEA studies: ''Although implicit in the IEA researchers' conceptualization of OTL was a belief that students should not be assessed on knowledge they had not been given an opportunity to learn'' (McDonnell, 1995, p. 306).

The link between OTL measures and equity became more explicit in the early 1990s, when OTL was addressed in policy debates on educational standards in the U.S. (Guiton & Oakes, 1995; McDonnell,

1995). In this context, it was assumed that ''if schools are to be held accountable for the equitable delivery of educational opportunities, the core of educational performance includes school and classroom process information'' (Wang, 1998, p. 137).

Beyond this debate, there is long-standing evidence that within countries, pronounced inequalities can be found in the way students are taught or exposed to the curriculum, especially in secondary education. ''Learning experiences are not only different for students from different SES families, they are also different for students in different ability or track classrooms'' (Wang, 1998, p. 138). Low-ability students, students attending vocational programs, or schools with an underprivileged social intake are exposed to an impoverished curriculum, both from a quantitative and a qualitative point of view (Hattie, 2009; Oakes, Gamoran, & Page, 1992).

OTL and achievement in reading in comparative assessments

The IEA-RLS results are fairly emblematic of the kind of issues or challenges generally raised when investigating the links between OTL and reading achievement in an international context.[1] A broad set of variables was collected from the teachers and organized into several OTL constructs, namely *comprehension instruction, emphasis on phonics, general emphasis on assessment, assessment of low-order skills, encouragement to read, taking students' interest into account, student-oriented reading, teacher-centered instruction,* and *reading aloud to students.* Lundberg (1994) analyzed the data for grade 4. At the descriptive level, he found that positive means of the standardized constructs (>.10) were mainly observed in non-European or Southern European countries; negative means were observed in Nordic and Western European countries, which may indicate a general cultural response bias (Yang et al., 2010). Typically, the lower-performing countries reported more positive average values on the OTL constructs. This led to negative correlations between OTL and achievement at the country level, even though, within each country, the correlation between OTL and reading achievement was positive. This phenomenon, known as the attitude-achievement paradox, is now well documented (Buckley, 2009; Van de Gaer et al., 2012). Lundberg (1994) applied a path analysis using the international pooled data set, controlling for socioeconomic background and resources, then introducing proximal teaching conditions (class size, instructional time), teacher's characteristics (education, experience, gender) and the six constructs for teaching reading. He found ''a rather small impact on student achievement, when factors related to home and community resources were partialled out'' (p. 185). Moreover, when the country was introduced as a control, very little variance was left to be explained. None of the teaching constructs had any significant effect. Replicating the same analysis at the national level, he found that very few significant correlations could be found between OTL/instructional practices and achievement.

Several explanations can be provided for these disappointing results: a measurement-related, a statistical and a methodological one. From a measurement point of view, OTL variables were self-reported measures, prone to common method bias. From a statistical point of view, controlling for students' social background and community resources may have led to an underestimation of the impact of OTL, because socioeconomic background and instructional practices are likely to have some joint variance. Finally, acquisition of reading skills is a cumulative process, and asking about

---

[1] In PIRLS, many OTL and instructional variables were also collected, but unfortunately the international reports only provide cross-tabulation statistics. Obviously, several papers use PIRLS data from one or a subset of countries and a selection of OTL or instructional variables (e.g. Lankes, 2004), but it is difficult to come up with a common synthesis of all these secondary analyses.

instructional practices or OTL with exclusive reference to the students' current teacher puts a strong methodological limitation on identifying efficient instructional practices in a cross-sectional study.

Conversely, experimental studies and meta-analyses provide long-term evidence that some instructional strategies significantly enhance reading comprehension (McNamara, 2007; Paris, Wasik, & Turner, 1991; Pressley, 2000; Rosenshine & Meister, 1997). On the basis of this literature, stronger links between OTL and achievement could be expected in comparative studies like PISA or PIRLS, if effective procedures aimed at overcoming some of the aforementioned limitations in OTL measurement were implemented.

## Aims of the study and hypotheses

Despite the fact that OTL is one of the best malleable predictors of achievement, use of OTL in international studies assessing reading literacy has been less systematic than for well-defined domains like mathematics and science. Moreover, when OTL data have been collected and analyzed, as in the IEA-RLS (Lundberg, 1994), results have been inconsistent and the attitude-achievement paradox has been observed (Van de Gaer et al., 2012). Common method or response-style bias was therefore suspected in OTL measures. With this in mind, we used a different methodological approach in order to overcome the common method bias. We indeed hypothesized that between- and within-country differences in response styles concealed substantial correlations between OTL and reading achievement, and that stronger associations would show up if we controlled for the common method bias linked to Likert scales.

After having applied this approach to improve the measurement of OTL, the present study investigated three research questions.

1. The first research question is related to equity. To what extent are students in each of the OECD countries offered similar or different opportunities to enhance their reading skills according to the school they are enrolled in? As many education systems with different structures take part in the PISA assessment, this is a unique opportunity to investigate whether education systems vary in the way they provide equal OTL in reading to all students.
2. The second research question explores the links between OTL and reading proficiency. To what extent can OTL in reading as they were measured in PISA explain reading proficiency within countries? How much variance of reading proficiency can be explained by OTL in reading uniquely or jointly with the socioeconomic composition of the school?
3. The third research question explores the extent to which the proportion of between-school variance in reading can be explained by OTL in reading, and whether this proportion of variance explained by OTL varies from one education system to another.

Question 1: on the basis of the literature on differentiation and adaptivity of the curriculum (Klieme & Warwas, 2011; Oakes et al., 1992; Wang, 1998), moderate to strong variance of OTL between schools was expected (Hypothesis 1a). More precisely, a stronger variance of OTL in reading was expected in differentiated education systems, whereas a more limited variance was expected in the more comprehensive ones (Hypothesis 1b). In this article, the term ''differentiated'' applies to education systems in which students are separated into different schools or tracks (academic, technical or vocational) according to their past performance, and in which the age of the first selection comes early, after primary education or before the end of lower secondary education (Mons, 2007). ''Comprehensive'' education refers to the opposite policy, in which students are not

sorted into different schools, tracks or classes according to their academic abilities before reaching the age of 16.

Question 2: drawing on the OTL literature and on what was observed in previous comparatives studies on reading literacy, moderate to robust correlations between OTL in reading and reading achievement were expected (Hypothesis 2a). Moreover, the more cognitively challenging texts and reading tasks (reading fiction texts, explaining the cause of events, the characters' behaviors, and the purpose of a text) were expected to show stronger positive correlations with reading proficiency than the less challenging ones *(reading texts with diagrams and maps, tables or graphs, finding information in a graph, and a diagram or a table)* (Hypothesis 2b).

Question 3: the proportion of between-school variance explained by OTL measures – uniquely or jointly with school social intake – was expected to be quite substantial (Hypothesis 3a), and more substantial in differentiated education systems than in more comprehensive education systems (Hypothesis 3b), because of the expected higher variance of OTL between schools/tracks (see Hypothesis 1b) in differentiated education systems and of the well-known robust link between academic and social segregation (Monseur & Lafontaine, 2009, 2012).

## Method

### Database

PISA 2009 (OECD, 2010a) assessed student achievement in reading as a major domain and mathematics and science as minor domains in 34 OECD and 21 partner countries. For the present study, only the OECD countries were analyzed. Representative student samples were obtained in each participating country according to a two-stage design. In each country, at least 150 schools were selected, with a probability proportional to their size. Simple random samples of thirty-five 15-year-old students per school were selected across classes, tracks and grades, to obtain an age-based sample of about 4500 students in each country.

### Variables

### Reading achievement

Reading achievement in PISA 2009 was measured as part of a 2-h testing session. Several types and formats of texts were used – continuous (organized in sentences and paragraphs, it could be descriptive, narrative, expository, argumentative and instructive texts) and non-continuous ones (or documents organized in lists, examples are tables, graphs, schedules, indexes and forms) (OECD, 2009, pp. 30–33). Moreover, three different aspects of reading were measured: retrieving information, interpreting the text, and evaluating and reflecting on the content and form of the text. Cognitive data were scaled according to a one-parameter logistic IRT model and students' level of reading proficiency was estimated by using plausible values. Results were reported on a combined scale, three per-aspect scales and two per-text-type subscales (continuous vs non-continuous).

### OTL in reading

In PISA 2009, 17 OTL items were used focusing on the reading materials and activities to which students were exposed in their classes[2] or during their homework. These items were included at the

---

[2] The question was asked in general (''in your classes''), no reference to specific or disciplinary classes.

end of the cognitive booklets, not in the questionnaires. According to Shanahan and Shanahan's typology (2008), the PISA OTL can be considered as mainly capturing ''intermediate literacy'', namely ''literacy skills common to many tasks, including generic comprehension strategies'' (p. 44) and some of them ''disciplinary literacy'', more specialized in literature. Eight items asked the students how often, during the last month, different types of texts were used in classes: Use of Fiction, of Poetry, of Texts that include diagrams or maps, of Texts that include tables or graphs, of Newspaper reports and magazine articles, of Instructions or manuals telling you how to make or do something, of Information texts about writers or books, of Advertising material (e.g., advertisements in magazines, posters). Nine items asked how often they were required to perform different types of tasks when reading these texts: *Explain the cause of events in a text, Explain the way characters behave in a text, Explain the purpose of a text, Learn about the life of the writer, Find information from a graph, diagram or table, Describe the way the information in a table or graph is organized, Explain the connexion between different parts of a text (e.g., between a written part and a graph), Memorize a text by heart, Learn about the place of a text in the history of literature.* For the present study, 11 out of the 17 original items were kept. The items removed from the analyses are: *reading information texts about writers or books, reading poetry, learning about the life of the writer, memorizing a text by heart, learning about the place of a text in the history of literature and Describe the way the information in a table or graph is organized.* Five out of these items target processes or content related to the ''disciplinary literacy'' (literature). The emphasis on literature in language of instruction classes largely varies from country to country. Consequently, such items positively correlated with achievement in some countries, especially in some Asian countries, and negatively in other countries. These items appeared to be problematic and were a source of instability in the Multi Group Confirmatory Analyses. The sixth item *Describe the way the information in a table or graph is organized* loaded on two different factors in the factorial analysis and was also a source of instability.

Background variables

The socioeconomic background of the student was measured via the PISA economic, social and cultural status index (ESCS). This index includes several components: home possessions (including cultural and educational resources); books at home; highest parental occupation and highest parental education (OECD, 2012, p. 312).

Statistical analyses

The 11 OTL items were modeled with a mixed coefficient multinomial model, as described by Adams, Wilson, and Wang (1997) and implemented by Conquest1 software (Wu, Adams, & Wilson, 1997). This mixed model describes items using a fixed set of unknown parameters, while student outcome level is a random effect (Adams, 2002). More precisely, the 11 OTL items were scaled according to a three-dimensional within-item partial credit model, with all OTL items loading on the first dimension and some loading specifically on dimension 2 or on dimension 3. This conditional model requires a population model defined by a set of repressors, usually denoted conditioning variables. In this study, several variables were used as conditioning variables: schools as dummies, reading performance, the PISA ESCS index, gender, the track attended by the student and a few indices such as reading enjoyment, teacher stimulation for reading engagement, use of structuring and scaffolding strategies, memorization, understanding and remembering, and elaboration. Five plausible values were drawn for each student per OTL dimension.

In order to estimate how OTL variables were distributed between schools (research question 1), intraclass correlations (ICC) of the OTL dimensions were computed. To address research question 2, correlations between OTL dimensions and reading achievement at the school level were computed (using the overall reading scale and the per-text format and per-aspect subscale). It was decided to model the impact of OTL at the school, but not at the student level. In comprehensive education systems, the within-school OTL variability mainly reflects measurement errors or individualization practices, while in highly differentiated systems it may also reflect the use of differentiated curricula within schools or even within-school tracking. Consequently, as the PISA sampling design does not allow modeling at the class level, the student and class level effects would have been confused. Due to these difficulties in interpreting within-school variation in OTL, and also because OTL is mainly a school or a class level effect, it would have been misleading to compare the impact at the student level between different education systems.

Finally, for research question 3, multilevel analyses (SAS 9.3) were used to test the amount of variance in reading achievement explained by the two OTL dimensions and the students' socioeconomic status (ESCS). Four[3] models were used separately for each country: model 1, an empty model, to estimate the between- and within-school components of the total variance in reading achievement; model 2, with the two school-level OTL dimensions as predictors; model 3, with the school aggregate of the PISA economic, social and cultural status index (ESCS) as predictor; and model 4, with all predictors from models 2 and 3, to estimate the unique and joint contribution of the OTL and school social intake.[4] Technical notes and country estimates of fixed and random parameters are provided in Appendix.

## Results

Modeling of the OTL

To overcome the common method bias of OTL (measured by Likert scales), a multidimensional within-item IRT approach was implemented, aimed at disentangling the target OTL from the common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Several models were tested; the validity and cross-cultural stability of the solutions are extensively described in Monseur, Baye, Vieluf, and Lafontaine (in preparation).

The rationale for allocating the items on the three dimensions was as follows. All the OTL items loaded on dimension 1, according to the results of factorial analysis before rotation (one factor); this dimension 1 was meant to capture some common OTL dimension and/or common method bias linked to Likert scales (acquiescence, disacquiescence, extremity scoring, intermediate response style, social desirability and/or ''satisficing'' behavior). Conversely, dimensions 2 and 3 were meant to represent the target constructs, namely OTL in reading. For these two target dimensions, drawing on what is known about the effectiveness of reading instruction strategies (McNamara, 2007; Pressley,

---

[3] Students in PISA are randomly sampled within schools. Unfortunately, this precludes the use of a three-level model (student, class and school), which would be the most adequate solution for OTL variables (Martinez, 2012).

[4] SES measures were not introduced at the student level as in highly differentiated education systems, this variable would have reduced substantially the between-school variance while in non-differentiated ones, the reduction would have been quite smaller. It would have therefore jeopardized the comparability of the results at the school level and, more importantly, it would have led to largely underestimate the joined variance at the school level between OTL and school SES.

2000; Rosenshine & Meister, 1997), OTL items were selected according to major features of the PISA framework (OECD, 2009), namely text format (continuous vs non-continuous) and aspect (retrieving information, interpreting and reflecting).

The three items allocated to dimension 2 were reading of non-continuous texts (texts with diagrams and maps, texts with tables and graphs) and one ''retrieving'' item in non-continuous texts, quite a straightforward task (finding information in a graph, a table or a diagram). The four items allocated to dimension 3 were reading fiction (novels, short stories) and three ''interpreting'' or ''reflecting'' items also related to fiction texts (explaining the cause of events in a text, explaining the way characters behave in a text and explaining the purpose of a text).

Cross-cultural invariance of the three dimensions was analyzed by means of multiple group confirmatory factor analysis (MGCFA) (Van de Vijver & Poortinga, 1982). The findings support the cross-national validity of the OTL dimensions and, to a certain degree, also its cross-cultural comparability. The criteria for configural and metric invariance were met, but not those for scalar invariance; this means that cross-national comparisons of correlations between OTL dimensions and reading achievement are valid, whereas mean scores on the OTL dimensions should not be compared between countries (for a more detailed description, see Monseur, Baye, Vieluf, and Lafontaine, in preparation).[5]

How equally is OTL in reading distributed between schools?

In order to estimate whether dimensions of OTL in reading are evenly distributed between schools, ICCs of the OTL dimensions were computed in each country. The ICC of the common method dimension was also computed. Results are displayed in Table 1.

The between-school variance of the common method factor was low in almost every country (0.17 on average among OECD countries, ranging from 0.08 in Finland to 0.34 in Italy). Conversely, the between-school variance of the two target OTL dimensions was higher on average; 26% of the total variance of dimension 2 and 31% of dimension 3 occurred between schools. The range was higher for target dimensions than for the common method dimension, from 0.14 to 0.48 for the dimension *non-continuous texts and retrieving information*, and from 0.14 to 0.60 for *fiction texts and interpreting/reflecting on fiction texts.* With few exceptions, the education systems in which a higher value of the ICC of the OTL dimensions was observed were mainly the differentiated ones (e.g. Austria, Belgium, Czech Republic, Germany, Hungary, Luxemburg, Slovakia, the Netherlands).

---

[5] As is common in cross-national studies (e.g. Cheung & Rensvold, 2002; Davidov, 2008; Steenkamp & Baumgartner, 1998), three nested MGCFA models were used to test the cross-cultural invariance of these dimensions. In the MGCFA models, one gradually tests different levels of invariance and compares the fit of models with different restrictions on the model parameters to determine what level of invariance can be assumed. The most basic level is configural invariance, which indicates that the same items form equivalent OTL factors across all participating countries – a precondition for all further cross-national comparisons. The second level of invariance, metric invariance, indicates that all factor loadings of the two OTL scales are equivalent across countries. This level is needed to make valid cross-national comparisons of intra-cultural score differences and of correlations with third constructs. Cross-national comparisons of mean scores additionally require so-called scalar equivalence, or equivalence of item-intercepts.

Table 1

*Per country intraclass correlations of the OTL and common method dimensions.*

| Country | Common method | OTL non-continuous | OTL fiction |
|---------|---------------|--------------------|-------------|
| AUS | 0.12 | 0.22 | 0.31 |
| AUT | 0.25 | 0.35 | 0.60 |
| BEL | 0.23 | 0.32 | 0.41 |
| CAN | 0.14 | 0.24 | 0.21 |
| CHE | 0.20 | 0.28 | 0.43 |
| CHL | 0.14 | 0.24 | 0.23 |
| CZE | 0.23 | 0.40 | 0.44 |
| DEU | 0.22 | 0.30 | 0.41 |
| DNK | 0.21 | 0.23 | 0.26 |
| ESP | 0.12 | 0.17 | 0.21 |
| EST | 0.15 | 0.26 | 0.28 |
| FIN | 0.08 | 0.21 | 0.29 |
| FRA | 0.18 | 0.27 | 0.30 |
| GBR | 0.18 | 0.22 | 0.25 |
| GRC | 0.16 | 0.14 | 0.17 |
| HUN | 0.28 | 0.34 | 0.37 |
| IRL | 0.15 | 0.15 | 0.14 |
| ISL | 0.12 | 0.26 | 0.23 |
| ISR | 0.20 | 0.26 | 0.37 |
| ITA | 0.34 | 0.20 | 0.37 |
| JPN | 0.10 | 0.18 | 0.30 |
| KOR | 0.10 | 0.24 | 0.24 |
| LUX | 0.13 | 0.40 | 0.34 |
| MEX | 0.16 | 0.26 | 0.45 |
| NLD | 0.20 | 0.35 | 0.44 |
| NOR | 0.17 | 0.28 | 0.34 |
| NZL | 0.12 | 0.18 | 0.26 |
| POL | 0.10 | 0.22 | 0.18 |
| PRT | 0.14 | 0.13 | 0.24 |
| SVK | 0.27 | 0.36 | 0.27 |
| SVN | 0.18 | 0.47 | 0.50 |
| SWE | 0.14 | 0.30 | 0.31 |
| TUR | 0.12 | 0.19 | 0.17 |
| USA | 0.15 | 0.19 | 0.26 |
| OECD average | 0.17 | 0.26 | 0.31 |

To what extent are the OTL dimensions related to reading achievement?

Correlations at the school level between the common method factor and the two OTL dimensions on the one hand and reading achievement (overall scale) on the other hand are displayed in Table 2.

The common method dimension was quite strongly negatively related to reading achievement in almost all countries (except in Finland and Korea, where the correlation was positive, but very low), while OTL dimensions showed quite robust positive correlations with achievement on average among the OECD countries.

Table 2

*Correlations at the school level between OTL and the common method dimensions on the one hand, reading achievement on the other hand.[a]*

| Country | Common method | OTL non-continuous | OTL fiction |
|---|---|---|---|
| AUS | -0.21 (0.086) | 0.41 (0.070) | 0.57 (0.045) |
| AUT | -0.60 (0.047) | 0.58 (0.037) | 0.70 (0.030) |
| BEL | -0.63 (0.040) | 0.49 (0.044) | 0.61 (0.038) |
| CAN | -0.10 (0.039) | 0.38 (0.038) | 0.33 (0.042) |
| CHE | -0.35 (0.069) | 0.36 (0.105) | 0.61 (0.056) |
| CHL | -0.33 (0.071) | 0.36 (0.078) | 0.38 (0.067) |
| CZE | -0.55 (0.047) | 0.49 (0.041) | 0.66 (0.041) |
| DEU | -0.42 (0.044) | 0.42 (0.041) | 0.61 (0.030) |
| DNK | -0.04 (0.063) | 0.25 (0.062) | 0.27 (0.076) |
| ESP | -0.36 (0.044) | 0.27 (0.049) | 0.50 (0.039) |
| EST | -0.30 (0.058) | 0.28 (0.056) | 0.25 (0.056) |
| FIN | 0.04 (0.083) | 0.28 (0.072) | 0.18 (0.096) |
| FRA | -0.51 (0.057) | 0.73 (0.032) | 0.71 (0.036) |
| GBR | -0.37 (0.054) | 0.50 (0.048) | 0.46 (0.069) |
| GRC | -0.31 (0.088) | 0.17 (0.080) | 0.66 (0.060) |
| HUN | -0.76 (0.038) | 0.67 (0.034) | 0.70 (0.043) |
| IRL | -0.29 (0.083) | 0.29 (0.076) | 0.22 (0.084) |
| ISL | -0.21 (0.028) | 0.34 (0.027) | 0.34 (0.037) |
| ISR | -0.55 (0.051) | 0.38 (0.066) | 0.28 (0.065) |
| ITA | -0.59 (0.023) | 0.14 (0.036) | 0.56 (0.024) |
| JPN | -0.33 (0.069) | 0.49 (0.046) | 0.65 (0.043) |
| KOR | 0.17 (0.116) | 0.68 (0.059) | 0.71 (0.057) |
| LUX | -0.69 (0.024) | 0.75 (0.011) | 0.89 (0.008) |
| MEX | -0.03 (0.042) | 0.13 (0.042) | 0.23 (0.040) |
| NLD | -0.59 (0.052) | 0.74 (0.032) | 0.72 (0.035) |
| NOR | -0.25 (0.064) | 0.38 (0.059) | 0.42 (0.056) |
| NZL | -0.58 (0.061) | 0.58 (0.057) | 0.50 (0.060) |
| POL | -0.53 (0.054) | 0.49 (0.063) | 0.58 (0.056) |
| PRT | -0.45 (0.065) | 0.47 (0.052) | 0.55 (0.055) |
| SVK | -0.62 (0.042) | 0.50 (0.054) | 0.53 (0.052) |
| SVN | -0.59 (0.014) | 0.74 (0.010) | 0.79 (0.008) |
| SWE | -0.02 (0.082) | 0.18 (0.071) | 0.25 (0.063) |
| TUR | -0.37 (0.069) | 0.30 (0.069) | 0.57 (0.049) |
| USA | -0.53 (0.067) | 0.44 (0.061) | 0.52 (0.072) |
| OECD average | -0.38 | 0.43 | 0.52 |

[a]Standard errors are shown in parentheses.

Correlations with the per-aspect and per-text format sub-scales were also computed. Very slight differences in correlations were observed by subscales. The OTL dimension *reading fiction, interpreting and evaluating fiction texts* showed, on average among OECD countries, very similar correlations with the aspects Interpreting (0.52), Reflecting/Evaluating (0.50) and Retrieving (0.49); the same was observed for the OTL dimension *reading non-continuous texts* (0.42, 0.43 and 0.42 respectively). Furthermore, no difference was observed for the correlation between OTL dimension

*reading non-continuous texts* and the two per-text-type subscales (0.43 and 0.42 respectively). The dimension *fiction* showed a somewhat stronger correlation with continuous (0.52) than with non-continuous texts (0.49). In addition, the OTL dimension *fiction* had a stronger link with both text format subscales (0.52 for continuous and 0.49 for non-continuous scales) than the OTL dimension *non-continuous texts* (0.43 for continuous and 0.42 for non-continuous scales). These slight variations in correlations are not surprising, as the per-aspect and per-text format subscales are themselves very highly correlated (around or above 0.90).

<u>How much of the between-school variance is explained by OTL?</u>

Table 3 presents the results of the multilevel regression analyses. It will be recalled that the dependent variable is student performance in reading and the independent variables are the school social intake (i.e. the school's mean student economic, social and cultural index) and the school aggregate of the two OTL factors. No student-level independent variables were included in the model.

As shown in column 2, on average, at the OECD level, 36% of the total variance in reading achievement lay between schools (model 1). As usual in PISA (OECD, 2013; Monseur and Lafontaine, 2009, 2012), large differences were observed between comprehensive education systems with low between-school variance in reading scores – the Scandinavian countries (Denmark, Finland, Iceland, Norway, Sweden), and a few other countries such as Poland, Canada and Spain – and differentiated education systems, in which a large amount of the variance in reading scores lies between schools – such as Austria, Belgium, Germany, Hungary, Italy, Luxembourg, the Netherlands or Slovakia.

Columns 3–6 present the percentages of variance explained at the school level by the school social intake and by the two OTL factors aggregated at the school level. Column 3 presents the percentage of school variance that is explained by all independent variables included in the model. At the OECD level, on average, the school social intake and the two OTL factors explain slightly more than two-thirds (precisely 68%) of the school variance. However, this total explained variance at the school level varies considerably. For instance, in Belgium, Hungary and Slovenia, 80% or more of the school variance is explained, while in Finland and Iceland, only about one-third of the school variance is explained.

However, school social intake and school OTL factors are obviously not independent. Schools attended by many students from an advantaged family background usually provide more opportunities to learn than do schools attended by more students from a disadvantaged family background. Additional analyses (see columns 4–6) took into account this potential overlap between school predictors by breaking down the total explained variance into three components:

1. The percentage of school variance that is uniquely explained by OTL (column 4). This represents the increase in the percentage of explained variance when OTL predictors are introduced into the model, in comparison with a model that has only the school social intake as a predictor. It also represents the percentage of school variance explained by OTL when school social intake is controlled for, i.e. if all schools had the same social intakes.
2. The percentage of school variance that is uniquely explained by the school social intake (column 5). Like the previous column, this represents the percentage of incremental explained variance linked to the introduction of school social intake.

3. The percentage of school variance explained jointly by the school social intake and the two OTL factors (column 6). This percentage reflects the overlap between the predictors in explaining the school average performance in reading, corresponding to the fact that OTL are not independent from school social intake.

Table 3

*Explained variance (in percentages) in reading performance at the school level.*

| CNT | Reading ICC | % Between-school variance explained | | | |
|---|---|---|---|---|---|
| | | Total | Unique OTL | Unique SES | Joint OTL + SES |
| AUS | 0.21 | 70 | 2 | 38 | 30 |
| AUT | 0.57 | 72 | 14 | 25 | 34 |
| BEL | 0.56 | 80 | 4 | 32 | 44 |
| CAN | 0.18 | 51 | 6 | 31 | 15 |
| CHE | 0.32 | 63 | 5 | 32 | 26 |
| CHL | 0.51 | 70 | 5 | 44 | 21 |
| CZE | 0.53 | 78 | 9 | 28 | 41 |
| DEU | 0.61 | 78 | 3 | 34 | 41 |
| DNK | 0.16 | 73 | 4 | 54 | 14 |
| ESP | 0.18 | 62 | 8 | 35 | 19 |
| EST | 0.22 | 54 | 6 | 43 | 6 |
| FIN | 0.07 | 32 | 9 | 19 | 4 |
| FRA | 0.58 | 77 | 17 | 21 | 39 |
| GBR | 0.20 | 74 | 3 | 39 | 32 |
| GRC | 0.42 | 68 | 21 | 27 | 20 |
| HUN | 0.69 | 84 | 11 | 26 | 47 |
| IRL | 0.21 | 60 | 3 | 46 | 10 |
| ISL | 0.14 | 36 | 13 | 11 | 12 |
| ISR | 0.48 | 58 | 5 | 38 | 15 |
| ITA | 0.54 | 60 | 6 | 27 | 27 |
| JPN | 0.49 | 67 | 5 | 21 | 40 |
| KOR | 0.32 | 76 | 18 | 14 | 43 |
| LUX | 0.34 | 84 | 11 | 19 | 53 |
| MEX | 0.45 | 50 | 0 | 49 | 2 |
| NLD | 0.62 | 74 | 19 | 13 | 43 |
| NOR | 0.10 | 55 | 16 | 23 | 16 |
| NZL | 0.21 | 76 | 9 | 31 | 36 |
| POL | 0.16 | 76 | 14 | 21 | 41 |
| PRT | 0.31 | 68 | 5 | 34 | 29 |
| SVK | 0.44 | 64 | 4 | 36 | 24 |
| SVN | 0.63 | 82 | 19 | 13 | 51 |
| SWE | 0.15 | 70 | 4 | 59 | 6 |
| TUR | 0.56 | 79 | 7 | 42 | 30 |
| USA | 0.24 | 75 | 4 | 40 | 31 |
| OECD average | 0.36 | 68 | 8 | 31 | 28 |

On average among OECD countries, the impact of OTL factors (8%) is negligible in comparison with the impact of school social intake (31%). In other words, if schools had a similar social intake, then OTL factors would explain only 8% of the school differences in reading performance. However, as illustrated by the joined variance, school social intake and school OTL are closely correlated, as on average, 28% of the school variance in reading performance is jointly explained by these predictors. The joined variance is particularly high in differentiated education systems like Belgium, the Czech Republic, Germany, Luxembourg and Slovenia.

## Discussion

Due to disappointing results in the past, researchers have been inclined to conclude that measuring OTL in reading in international large-scale assessments was not worthwhile, and were prone to attribute this lack of positive relationship between OTL and achievement in reading to the fact that reading was not taught as a specific subject, especially in secondary education.

The results of the multidimensional within-item IRT applied to model the OTL items support the hypothesis that answers to the OTL items were affected by a common method bias. Dimension 1, to which all items were allocated, was significantly and negatively related to achievement in almost all countries (except Finland and Korea). Thus, paradoxically, within countries, students enrolled in schools performing less well reported more frequent exposure to presumably beneficial OTL. In addition, at the country level, the highest values on dimension 1 were observed in Chile, Mexico and Turkey – the three OECD countries performing least well in PISA. These paradoxical results make sense if dimension 1 is interpreted as a common method dimension, capturing not only some common OTL dimension, but also response styles, namely the tendency to overclaim or be acquiescent whatever the item. Indeed, this tendency is more frequent among low-performing individuals or in low-performing or low-GDP countries, according to the findings from cross-cultural psychology (Johnson et al., 2005).

Since common method bias within each of the OTL items was captured by dimension 1, dimensions 2 and 3 were at least partially ''purged'' of the response style bias, and were therefore able to better grasp the target construct of OTL in reading, namely exposure to different text formats and reading tasks related to them. It was thus possible to explore our research questions drawing on a more valid and targeted measurement of OTL in reading than in previous comparative studies.

Our first research question was related to equity in OTL exposure. To what extent are students offered equal opportunities to enhance their reading skills whatever the school they are enrolled in; and does this differential exposure to OTL by school vary according to the structure of the education system (differentiated vs comprehensive structure in lower secondary education)? First of all, the between-school variance in the two target OTL dimensions was quite high on average, almost as high as the between-school variance in reading achievement itself. Students were definitely exposed to beneficial OTL in reading more often in some schools than others; Hypothesis 1a is thus confirmed. As hypothesized (Hypothesis 1b), the between-school variance in OTL was higher in differentiated systems (especially in Austria, Belgium, the Czech Republic, Switzerland, Germany, Hungary, Luxembourg, Mexico, the Netherlands, Slovakia) than in comprehensive ones (Canada, Finland, Greece, Spain, Ireland, Poland and Portugal).

Quite robust correlations between OTL in reading and reading achievement were observed at the student[6] and at the school level – far more robust than when a classical factorial analysis[7] is used to build OTL constructs (for more details, see Monseur, Baye, Vieluf, and Lafontaine, in preparation).[8] Hypothesis 2a is thus confirmed. Furthermore, as we hypothesized, the OTL dimension *reading fiction* and *interpreting/reflecting on fiction texts*, grouping more cognitively challenging texts and tasks (interpret and reflect versus retrieve information in non-continuous texts), had a stronger correlation with reading achievement than the OTL dimension *reading non-continuous texts* and *retrieving information in non-continuous texts*. Hypothesis 2b is thus confirmed too.

As hypothesized, the proportion of the between-school variance explained by the OTL dimensions uniquely or jointly with school social intake was quite substantial (Hypothesis 3a) – 68% on average among OECD countries. Moreover, the amount of variance explained jointly by OTL and school social intake was on average higher than the variance explained uniquely by social intake. This is why, in previous studies, especially Lundberg (1994), OTL variables have barely explained any variance, once students' socioeconomic background has been partialled out. This way of analyzing data in the past has led to the potential impact of OTL in reading being underestimated, given the importance of the joint variance between OTL and socioeconomic status.

In addition, correlations between the proportion of between-school variance in reading achievement explained either uniquely by OTL or jointly by OTL and school social intake, and an indicator of differentiation of the education system structure (namely the proportion of 15-year-olds attending a non-academic track) were positive and significant. Hypothesis 3b is thus confirmed. This suggests that variation in curriculum was more often observed in differentiated education systems, which makes perfect sense. By definition, when students are grouped in schools or tracks according to their abilities, it is to allow for differentiation (Oakes et al., 1992) and macro-adaptivity of the curriculum (Klieme & Warwas, 2011). Moreover, our results showed that especially the joint variance OTL/school social intake was much stronger in differentiated education systems. In differentiated education systems, curricular variations in OTL were obviously more closely linked to the socioeconomic composition of the schools. Those results help to better understand why differentiated education systems are less equitable than comprehensive ones: exposure to reading OTL is different from school to school, and this differential exposure is linked to the socioeconomic composition of the school. Students attending schools with a less privileged social intake are more exposed to impoverished OTL in reading (less demanding texts and tasks). As OTL potentially has a strong impact on achievement, this differential exposure to an influential variable leads to a larger gap between low and high achievers and between more and less privileged students in differentiated education systems than in comprehensive ones (OECD, 2010b; Monseur & Lafontaine, 2012).

Limitations and further research

One serious limitation of this study is that no data at the classroom level are available, due to the way students are sampled in PISA. Therefore, the analyses could only be performed at the student

---

[6] Figures for the correlations at the student level are not provided in the paper, but on average the dimension *Non continuous* had a 0.32 and the dimension Fiction a 0.40 correlation with reading achievement.

[7] This CFA was attempted in previous steps, as well as a MGCFA showing that configural and metric invariance of the four factors was present, but scalar invariance was not.

[8] Two factors were close to dimensions 2 and 3 of the IRT model: ''Use of non-continuous texts'' and ''Interpretation of fiction texts''. The correlations with reading achievement at the student level were 0.07 and 0.25, respectively. In the CFA, common method and target constructs were confounded.

and at the school levels. Martinez (2012), in a methodological study about reading in U.S. primary schools, has shown that the classroom is the more appropriate level to analyze OTL variables. ''Ignoring the classroom level distorts the school and student level variance components'' (Martinez, 2012,

p. 308; see also Noortgate, Opdenakker, & Onghena, 2005). For secondary education, the direction of this bias is less well documented. Another potential limitation is that OTL was reported by students, which gives a limited perspective about teachers' instructional practices. However, there is evidence that the OTL reported by students is close to that reported by teachers. Martinez and Goldschmidt (2003) found similar factorial structures and a broad overlap between teachers and students' reports on OTL. Finally, PISA data are cross-sectional, so no causal inferences are allowed regarding the impact of OTL in reading on achievement.

Another limitation is linked to the items that were included in PISA to measure the OTL. Even if the items are closely related to the aspects and the text format – two main dimensions of the PISA reading framework – and measure content coverage or emphasis specifically linked to the PISA assessment, other potentially important and relevant aspects of the reading literacy curriculum in secondary education were not included. By construct, the OTL PISA items were more focused on ''intermediate'' or ''generic'' literacy than on ''disciplinary literacy'' (Shanahan & Shanahan, 2008), as is the PISA assessment itself. Reading literacy experts could argue that areas of content or disciplinary literacy were not enough taken into account in PISA. Even if discussing the limitations of PISA is out of scope of our study, it should be highlighted that the very fact that PISA deals with more than 70 education systems implies a huge heterogeneity of curricula and study programs between and within country; in addition, in some countries, 15 year-olds are still attending basic lower secondary education, while others are already in high school. In that context, it seems pragmatic to focus on ''common'' or ''generic'' reading literacy and OTL.

Conclusions and implications

Empirical studies using data on OTL in reading from large-scale international assessments are scarce, especially in secondary education. In addition, these studies are not recent (Grisay, 2008; Lundberg, 1994) and give quite disappointing results: only a small amount of the reading variance is explained by OTL in reading or other instructional variables. On this basis, researchers have tended to conclude that there is no point taking OTL in reading into account in large-scale international studies. The results of the present study lead to a more optimistic conclusion. Two reading OTL dimensions – firstly reading and interpreting fiction texts, and secondly reading non-continuous texts and retrieving information in this kind of text – showed positive and robust correlations with reading achievement. In the light of these results, there is no doubt that measuring OTL in reading in large-scale international assessments is not only worthwhile, but very important. The multidimensional within-item IRT model method used to overcome method bias proved to be promising; it could be applied to other domains, or to constructs measured using Likert-type scales which have proved to suffer from common method bias (Buckley, 2009). The new perspectives opened up by this approach are especially relevant in the context of large-scale international studies in which not only are individual response styles observed, but cross-cultural biases are also at play, leading to the attitude-achievement paradox.

A substantial amount of the variance in reading was explained by OTL in reading, either uniquely or jointly with school social intake. This overlap between curricular variation and school social intake

was the main component in many countries. Furthermore, the present study shows that exposure to OTL in reading varies from school to school, to a limited extent in comprehensive education systems, but to a strong extent in differentiated ones. In the latter, not only is a substantial per school variation in OTL observed, but these curricular variations are also more strongly related to the school's social composition. The joint variance between school social intake and OTL is possibly one of the main mechanisms that lead to an increase in academic and social inequalities and gaps between students and schools. By identifying the magnitude of the joint variance between OTL and school social intake in differentiated education systems, our study has shed some light on why differentiated education systems are less equitable than comprehensive ones (OECD, 2010b; Monseur & Lafontaine, 2009, 2012).

## Acknowledgments

## Appendix

The multilevel analyses, unlike the other analyses presented in this study, were conducted only on the first plausible values in reading and in the OTL measures.

Eqs. (1)–(4) respectively present the 4 models used to compute the percentages of explained variance (see Table 3).

(1) $Y_{ij} = \beta_{0j} + \epsilon_{ij}$
$\beta_{0j} = \gamma_{00} + U_{0j}$

(2) $Y_{ij} = \beta_{0j} + \epsilon_{ij}$
$\beta_{0j} = \gamma_{00} + \gamma_{01}(OTL2_j) + \gamma_{02}(OTL3_j) + U_{0j}$

(3) $Y_{ij} = \beta_{0j} + \epsilon_{ij}$
$\beta_{0j} = \gamma_{00} + \gamma_{01}(SES_j) + U_{0j}$

(4) $Y_{ij} = \beta_{0j} + \epsilon_{ij}$
$\beta_{0j} = \gamma_{00} + \gamma_{01}(OTL2_j) + \gamma_{02}(OTL3_j) + \gamma_{03}(SES_j) + U_{0j}$

Table A1 presents the random parameter estimates of Eq. (1) (i.e. the unconditional between-school and within-school variance estimates), Tables A2 and A3 respectively present the conditional random parameter estimates and the fixed parameter estimates of Eq. (4).

Table A1

*Unconditional between-school and within-school variance estimates of the student reading performance (see Eq. (1)).*

| Country | Between-school variance | | Within-school variance | |
|---|---|---|---|---|
| | Estimate | SE | Estimate | SE |
| AUS | 1873 | 161.8 | 7183 | 88.6 |
| AUT | 5770 | 532.7 | 4362 | 78.6 |
| BEL | 6043 | 542.8 | 4749 | 75.8 |
| CAN | 1455 | 90.1 | 6537 | 62.9 |
| CHE | 2575 | 199.4 | 5457 | 72.8 |
| CHL | 3887 | 433.2 | 3788 | 73.9 |
| CZE | 4899 | 472.7 | 4354 | 83.1 |
| DEU | 5925 | 578.6 | 3740 | 77.3 |
| DNK | 1076 | 127.9 | 5769 | 110.6 |
| ESP | 1312 | 82.3 | 5871 | 53.8 |
| EST | 1511 | 189.7 | 5276 | 114.0 |
| FIN | 519 | 80.8 | 6834 | 129.8 |
| FRA | 6335 | 725.1 | 4627 | 105.3 |
| GBR | 1684 | 142.2 | 6765 | 89.8 |
| GRC | 4026 | 457.5 | 5500 | 115.4 |
| HUN | 6336 | 683.5 | 2821 | 61.1 |
| IRL | 1741 | 242.1 | 6531 | 155.2 |
| ISL | 1239 | 226.1 | 7844 | 195.2 |
| ISR | 5994 | 666.0 | 6520 | 126.4 |
| ITA | 4693 | 217.1 | 3920 | 32.9 |
| JPN | 4812 | 518.2 | 4913 | 91.2 |
| KOR | 1915 | 234.7 | 4011 | 82.5 |
| LUX | 3551 | 823.0 | 6898 | 146.9 |
| MEX | 2888 | 114.5 | 3593 | 27.8 |
| NLD | 4650 | 497.7 | 2805 | 59.7 |
| NOR | 817 | 120.5 | 7351 | 158.6 |
| NZL | 2100 | 268.5 | 8124 | 174.7 |
| POL | 1289 | 176.6 | 6783 | 142.2 |
| PRT | 2329 | 250.6 | 5156 | 93.8 |
| SVK | 3344 | 367.3 | 4280 | 93.6 |
| SVN | 5121 | 427.6 | 2984 | 56.2 |
| SWE | 1422 | 192.4 | 7809 | 170.5 |
| TUR | 4095 | 471.7 | 3268 | 70.9 |
| USA | 2220 | 274.0 | 7012 | 140.4 |

## Table A2

*Conditional between-school and within-school variance estimates of the student reading performance (see Eq. (4)).*

| Country | Between-school variance | | Within-school variance | |
|---|---|---|---|---|
| | Estimate | SE | Estimate | SE |
| AUS | 561 | 58.4 | 7176 | 88.4 |
| AUT | 1612 | 176.2 | 4375 | 79.0 |
| BEL | 1210 | 122.7 | 4752 | 75.8 |
| CAN | 716 | 51.3 | 6537 | 62.9 |
| CHE | 949 | 89.0 | 5465 | 72.9 |
| CHL | 1150 | 144.8 | 3792 | 74.0 |
| CZE | 1076 | 118.0 | 4344 | 82.7 |
| DEU | 1299 | 142.3 | 3742 | 77.4 |
| DNK | 296 | 49.0 | 5731 | 109.2 |
| ESP | 493 | 37.7 | 5855 | 53.5 |
| EST | 693 | 103.8 | 5289 | 114.5 |
| FIN | 353 | 62.2 | 6828 | 129.7 |
| FRA | 1441 | 192.7 | 4643 | 105.9 |
| GBR | 441 | 48.7 | 6729 | 88.8 |
| GRC | 1294 | 166.3 | 5504 | 115.5 |
| HUN | 1012 | 126.5 | 2826 | 61.3 |
| IRL | 703 | 116.5 | 6556 | 156.3 |
| ISL | 791 | 164.1 | 7842 | 195.1 |
| ISR | 2516 | 293.3 | 6520 | 126.4 |
| ITA | 1877 | 93.8 | 3918 | 32.9 |
| JPN | 1604 | 184.6 | 4913 | 91.2 |
| KOR | 466 | 68.1 | 4008 | 82.4 |
| LUX | 576 | 164.9 | 6904 | 147.1 |
| MEX | 1431 | 61.3 | 3592 | 27.8 |
| NLD | 1190 | 138.4 | 2805 | 59.6 |
| NOR | 365 | 70.8 | 7336 | 158.1 |
| NZL | 499 | 91.9 | 8132 | 175.0 |
| POL | 310 | 61.8 | 6769 | 141.5 |
| PRT | 744 | 94.1 | 5157 | 93.8 |
| SVK | 1205 | 151.2 | 4288 | 93.9 |
| SVN | 919 | 94.8 | 2983 | 56.1 |
| SWE | 430 | 82.5 | 7809 | 170.3 |
| TUR | 869 | 111.6 | 3267 | 70.8 |
| USA | 553 | 89.4 | 7015 | 140.6 |

## Table A3

*Level 2 fixed parameter estimates (see Eq. (4)).*

| Country | Intercept | | School OTL Non-continuous | | School OTL Fiction | | School SES | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| AUS | 403 | 22.5 | 0.04 | 0.043 | 0.13 | 0.041 | 82.6 | 4.60 |
| AUT | 241 | 23.1 | 0.18 | 0.054 | 0.31 | 0.048 | 80.1 | 5.96 |
| BEL | 346 | 25.1 | -0.02 | 0.059 | 0.33 | 0.056 | 101.3 | 5.27 |
| CAN | 367 | 13.2 | 0.17 | 0.023 | 0.07 | 0.023 | 56.9 | 2.96 |
| CHE | 383 | 20.9 | 0.03 | 0.043 | 0.19 | 0.038 | 90.1 | 5.64 |
| CHL | 267 | 41.5 | 0.10 | 0.080 | 0.33 | 0.077 | 50.2 | 3.33 |
| CZE | 350 | 19.7 | 0.03 | 0.041 | 0.29 | 0.035 | 114.7 | 7.10 |
| DEU | 322 | 34.2 | 0.02 | 0.054 | 0.28 | 0.057 | 115.4 | 6.61 |
| DNK | 400 | 18.9 | 0.06 | 0.037 | 0.10 | 0.031 | 64.4 | 3.83 |
| ESP | 329 | 15.5 | 0.09 | 0.023 | 0.25 | 0.028 | 42.7 | 1.92 |
| EST | 363 | 33.1 | 0.22 | 0.070 | 0.03 | 0.060 | 63.5 | 6.05 |
| FIN | 415 | 25.5 | 0.16 | 0.045 | 0.06 | 0.053 | 32.7 | 6.19 |
| FRA | 142 | 37.2 | 0.39 | 0.090 | 0.28 | 0.083 | 85.1 | 7.73 |
| GBR | 346 | 20.9 | 0.11 | 0.036 | 0.14 | 0.040 | 83.0 | 4.23 |
| GRC | 277 | 37.7 | -0.22 | 0.082 | 0.58 | 0.058 | 55.5 | 5.02 |
| HUN | 213 | 30.4 | 0.20 | 0.063 | 0.36 | 0.056 | 69.0 | 4.38 |
| IRL | 376 | 40.0 | 0.12 | 0.073 | 0.12 | 0.077 | 64.7 | 6.10 |
| ISL | 320 | 37.1 | 0.06 | 0.075 | 0.29 | 0.080 | 25.4 | 7.88 |
| ISR | 270 | 50.1 | 0.20 | 0.095 | 0.22 | 0.084 | 110.1 | 9.09 |
| ITA | 336 | 17.1 | 0.03 | 0.033 | 0.25 | 0.024 | 76.1 | 3.05 |
| JPN | 352 | 35.3 | 0.12 | 0.085 | 0.22 | 0.066 | 112.6 | 10.90 |
| KOR | 340 | 22.6 | 0.20 | 0.061 | 0.24 | 0.075 | 45.9 | 5.56 |
| LUX | 294 | 58.6 | -0.17 | 0.100 | 0.49 | 0.108 | 68.9 | 11.53 |
| MEX | 451 | 15.7 | 0.00 | 0.028 | 0.07 | 0.035 | 42.1 | 1.19 |
| NLD | 97 | 35.3 | 0.37 | 0.084 | 0.43 | 0.084 | 67.0 | 7.41 |
| NOR | 344 | 21.3 | 0.10 | 0.037 | 0.16 | 0.045 | 54.0 | 7.20 |
| NZL | 312 | 34.0 | 0.30 | 0.067 | 0.10 | 0.066 | 75.3 | 6.67 |

| Country | Intercept | | School OTL Non-continuous | | School OTL Fiction | | School SES | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| POL | 317 | 26.6 | 0.16 | 0.047 | 0.21 | 0.043 | 40.5 | 4.39 |
| PRT | 361 | 32.8 | -0.01 | 0.066 | 0.28 | 0.064 | 51.2 | 3.92 |
| SVK | 396 | 24.7 | 0.03 | 0.051 | 0.15 | 0.051 | 90.9 | 7.39 |
| SVN | 192 | 19.0 | 0.13 | 0.044 | 0.44 | 0.038 | 68.1 | 5.04 |
| SWE | 349 | 33.9 | 0.11 | 0.058 | 0.14 | 0.060 | 81.1 | 5.84 |
| TUR | 382 | 32.8 | -0.10 | 0.060 | 0.34 | 0.050 | 57.3 | 3.44 |
| USA | 349 | 31.1 | 0.14 | 0.055 | 0.11 | 0.050 | 64.4 | 4.80 |

# References

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23.

Buckley, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006.* https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9, 233–255.* http://dx.doi.org/10.1207/S15328007SEM0902_5

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey. *Survey Research Methods*, 2, 33–46.

Dubet (2010). *Les places et les chances. Repenser la justice sociale*. Paris: Le Seuil.

Duru-Bellat (2009). *Le mérite contre la justice*. Paris: Presses de Sciences po.

EGREES (2005). Equity in European educational systems: A set of indicators. *European Educational Research Journal*, 4(2), 1–151.

Felouzis, G. (2014). *Les inégalités scolaires*. Paris: Presses Universitaires de France.

Floden, R. E. (2002). The measurement of opportunity-to-learn. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 231–267). Washington, DC: National Academy Press.

Grisay, A. (1984). Quels indicateurs pour quelle réduction des inégalités scolaires? *Revue de la Direction générale de l'organisation des études*, 9, 3–14.

Grisay, A. (2008). Opportunity to learn in grade 4 classes: Reading instruction. In Y. Zhang, T. N. Postlethwaite, & A. Grisay (Eds.), *A view inside primary schools: A World Education Indicators (WEI) cross-national study* (pp. 175–207). Montreal: UNESCO Institute of Statistics.

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

Harzing, A. W. K. 26 collaborators. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review*, 18(4), 417–432.

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

Johnson, T. J., Kulesa, P., Cho, Y., & Shavitt, S. (2005). The relation between culture and styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36(2), 264– 277.

Klieme, E., & Warwas, J. (2011). Konzepte der Individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805–818.

Kyllonen, P. C., & Bertling, J. P. (2014). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment data analysis*. London: Chapman and Hall/CRC.

Lankes, E.-M. (2004). Leseunterricht in der Grundschule. Unterschiede zwischen Lehrkräften im Internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 7, 551–568.

Leinhardt, G., Zigmond, N., & Cooley, W. W. (1981). Reading instruction and its effects. *American Educational Research Journal*, 18(3), 343–361.

Lundberg, I. (1994). The teaching of reading. In W. B. Elley (Ed.), *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems* (pp. 149–193). Pergamon.

Martinez, F. (2012). Consequences of omitting the classroom level in multi-level models of schooling: An illustration using opportunity-to-learn and reading achievement. *School Effectiveness and School Improvement*, 23(3), 305–320.

Martinez, F., & Goldschmidt, P. (2003). Comparing student and teacher reports on opportunity-to-learn and their relationship to student achievement: A multilevel analysis approach. *Paper presented at the Annual meeting of the American Educational Research Association.*

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

McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theories, interventions and technologies.* NJ: L. Erlbaum.

Meuret, D. (1999). Rawls, l'éducation et l'égalité des chances. In D. Meuret (Ed.), *La justice du système éducatif.* Bruxelles: De Boeck Université.

Mons, N. (2007). *Les nouvelles politiques éducatives*. Paris: PUF.

Monseur, C., & Lafontaine, D. (2009). L'organisation des systèmes éducatifs: quel impact sur l'efficacité et l'équité? In V. Dupriez & X. Dumay (Eds.), *L'efficacité en éducation, promesses et zones d'ombre* (pp. 141–165). Brussels: De Boeck.

Monseur, C., & Lafontaine, D. (2012). Structure des systèmes éducatifs et équité: un éclairage international. In M. Crahay (Ed.), *L'école peut-elle être juste et efficace* (pp. 185–221). Brussels: De Boeck.

Monseur, C., Baye, A., Magis, D., Vieluf, S., & Lafontaine, D. (in preparation). Towards a more valid measure of reading opportunities-to-learn in international large-scale assessments.

Muthén, B., Huang, L.-C., Jo, B., Khoo, S.-T., Goff, G. N., Novak, J. R., et al. (1995). Opportunity-to-learn effects on achievement: Analytical aspects. *Educational Evaluation and Policy Analysis*, 20(3), 371–403.

Noortgate, W., Opdenakker, M.-C., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement*, 16(3), 281–303.

Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570–608). New York: Macmillan.

OECD (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.

OECD (2010a). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science*. Paris: OECD.

OECD (2010b). *PISA 2009 results: Overcoming social background. Equity in learning opportunities and outcomes.* Paris: OECD.

OECD (2012). *PISA 2009 technical report*. Paris: OECD.

OECD (2013). *PISA 2012 results: What makes schools successful? Resources, policies and practices.* Paris: OECD.

Paris, S., Wasik, B., & Turner, J. (1991). The development of strategic readers. In R. Barr, L. Kamil, P. Mosenthal, & D. Pearson (Eds.), *Handbook of reading research* (Vol. I, pp. 609–640). NJ: L. Erlbaum.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 545–563). NJ: L. Erlbaum.

Rawls, J. (1971). *A theory of justice*. Oxford: University Press.

Rocereto, J. F., Puzakova, M., Anderson, R. E., & Kwak, H. (2011). The role of response formats on extreme response style: A case of Likert-type vs. semantic differential scales. In M. Sarstedt, M. Schwaiger, & C. R. Taylor (Eds.), *Measurement and research methods in international marketing (Advances in international marketing,* Vol. 22) (pp. 53–71). Emerald Group Publishing Limited.

Rosenshine, B., & Meister, C. (1997). Cognitive strategy instruction in reading. In A. Stahl & A. Hayes (Eds.), *Instructional models in reading* (pp. 85–107). NJ: L. Erlbaum.

Schmidt, W. H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research*. New York: Routledge.

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.

Sen, A. (1992). *Inequality re-examined*. Oxford: Oxford University Press.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content area literacy. *Harvard Educational Review*, 78(1), 40–60.

Steenkamp, J.-B.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90. http://dx.doi.org/10.1086/209528

Stevens, F. I. (1993). Applying an opportunity-to-learn conceptual framework to the investigation of teaching practices via secondary analyses of multiple-case study summary data. *Journal of Negro Education*, 62(3), 232–248.

Van de Gaer, E., Grisay, A., Schlutz, W., & Gebhardt, E. (2012). The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries. *Journal of Cross-cultural Psychology* http://jcc.sagepub.com/content/early/2012/02/08/0022022111428083

Van de Vijver, F., & Poortinga, Y. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology*, 13, 387–408.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis*, 20(3), 137–156.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *Conquest®: Multi-aspect test software (computer program manual).* Camberwell, VIC: Australian Council of Educational Research.

Yang, Y., Harkness, J. A., Chin, T.-Y., & Villar, A. (2010). Response styles and cultures. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, P.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional and multicultural contexts* (pp. 203–233). Hoboken: Wiley & Sons.

Dr. Dominique Lafontaine is a full professor of Educational Sciences at the University of Liège and director of the Center for Analysis of Systems and Practices in Education (aSPe). She has been involved in several large-scale assessment programs, at both national (IEA Reading Literacy, PIRLS 2006, PISA) and international level. She has been member of the PISA Reading Expert Group since 1998. In PISA, her interests have always been the non-cognitive aspects of reading (reading practices and attitudes, engagement, opportunity to learn in reading).

Dr. Ariane Baye is a professor at the University of Liège. She mainly lectures on diversity and prevention of early-school leaving and exclusion. She has been involved in PISA at both national and international level. Her research interests include equity, diversity, comparative education and reading.

Dr. Svenja Vieluf is a psychologist and post-doctoral researcher at the German Institute for International Educational Research (DIPF) in Frankfurt, Germany. She has been involved in instrument development and reporting for the OECD large scale surveys TALIS and PISA. Her research interests include teaching quality, multicultural education, comparative education and cross-cultural research methods.

Dr. Christian Monseur is a professor at the University of Liège. He mainly lectures on quantitative research methods, including sampling theory, measurement models and statistical analyses. He has been involved in large-scale assessment at both national and international level. He has been a

member of the IEA Technical Executive Group and of the Technical Expert Group of the IEA ICCS study, and is currently a member of the OECD PISA Technical Expert Group and of the OECD PIAAC survey.