Learning and Individual Differences 65 (2018) 50–64

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA

Jihyun Lee[a,*], Lazar Stankov[b,c]

[a] School of Education, The University of New South Wales, Sydney, Australia
[b] School of Psychology, The University of Sydney, Australia
[c] School of Psychology, The University of Southern Queensland, Australia

ARTICLE INFO

Keywords:
PISA
TIMSS
Academic achievement
Self-beliefs
Self-efficacy

ABSTRACT

We examined the predictability of non-cognitive variables for students' mathematics achievement, based on large-scale international databases of the TIMSS 2003, 2007, and 2011, and the PISA 2003 and 2012. We synthesized empirical evidence about 65 non-cognitive variables, which were categorized into 13 research domains of educational psychology—*affect, curriculum/content exposure, homework, learning and instructional time, motivation, personality traits, planned behavior, school climate, self-beliefs/social-cognitive theory, self-regulatory learning style/strategies, teacher behavior, value,* and *vocational interest*. Our analyses showed that a group of *self-beliefs* constructs, in particular, *self-efficacy* in PISA, *confidence* in TIMSS, and *educational aspiration,* in both TIMSS and PISA, were the best predictors of individual-level student achievement in mathematics. The present review supports the claim that students' projective judgements about their own ability and future selves are particularly important for their academic achievement. We discuss potential educational initiatives to maximize educational outcomes of students from diverse cultural and national backgrounds.

1. Introduction

Various social science disciplines, including economics (e.g., Heckman, Stixrud, & Urzua, 2006), education (Duckworth & Yeager, 2015; Lee & Shute, 2010), sociology (Bowles & Gintis, 1976; Farkas, 2003), and psychology (Richardson, Abraham, & Bond, 2012; Stankov, 2013) have devoted considerable research effort towards identifying which "non-cognitive" attributes are of relevance to students' academic achievement (Cunha, Heckman, & Schennach, 2010). The term "non-cognitive" is commonly used to refer to a broad range of personal attributes, skills and characteristics representing one's attitudinal, behavioral, emotional, motivational and other psychosocial dispositions. Sociologists have used the term "non-cognitive" as a catch-all phrase encompassing variables that are potentially important for academic achievement, but which are not measured by typical achievement or cognitive tests (Bowles & Gintis, 1976, p. 135; Farkas, 2003, p. 542). According to Duckworth and Yeager (2015), no satisfactory replacement label for the term "non-cognitive" has yet been proposed, in spite of recurring criticisms of the term itself in the current research literature. Non-cognitive constructs can be seen as: "(a) conceptually independent from cognitive ability, (b) generally accepted as beneficial to the student and to others in society, (c) relatively rank-order stable over time in the absence of exogenous forces, (d) potentially responsive to intervention, and (e) dependent on situational factors for their expression" (p. 239; see also Messick, 1979, 1996). It is generally believed that students' non-cognitive attributes are open to change through appropriate schooling and interventions (Duckworth & Yeager, 2015; Lee & Shute, 2010), while other environmental factors (e.g., family socio-economic status) are distal and removed from the direct influence of schooling. A core underlying assumption in much of this literature is that students themselves are the most important factor in the attainment of their educational and life outcomes.

In spite of the importance assigned to non-cognitive constructs, there have been no *conclusive* recommendations as to which non-cognitive attributes are likely to be most useful and have direct relevance to students' academic achievement. What we mean by *conclusive* is that research evidence should be systematically and comprehensively examined and synthesized based on a large number of variables examined in data from international samples so that the findings can be reasonably seen to apply to students from different cultural and national backgrounds. In the present paper, we aim to do so by documenting and integrating empirical evidence of predictive validity of students' non-cognitive attributes in relation to their academic achievement, using databases drawn from two widely-known, recent large-scale

* Corresponding author at: School of Education, UNSW-Sydney, Sydney, Australia.
E-mail address: jihyun.lee@unsw.edu.au (J. Lee).

https://doi.org/10.1016/j.lindif.2018.05.009
Received 7 June 2017; Received in revised form 6 May 2018; Accepted 12 May 2018
1041-6080/ © 2018 Elsevier Inc. All rights reserved.

international assessments: (a) the Trends in International Mathematics and Science Study (TIMSS) administered by the International Association for the Evaluation of Educational Achievement (IEA), and (b) the Programme for International Student Assessment (PISA) administered by the Organisation for Economic Co-operation and Development (OECD). We also compare our results with those of recent review papers that had similar research aims. Noteworthy examples are found in the reviews by Hattie (2009), Richardson et al. (2012), Stankov (2013), and Lee and Shute (2010). These are summarized in the next section.[1]

1.1. Four recent reviews of non-cognitive constructs and academic achievement

1.1.1. Hattie (2009)

Hattie's (2009) book, *Visible Learning*, presents a mega-analysis of the findings from > 800 meta-analytic studies. After evaluating an extensive number of non-cognitive constructs, he concludes that the four "best" student-level constructs in relation to academic achievement are: *engagement and motivation* (Cohen's $d = 0.48$), *self-concept* (Cohen's $d = 0.43$), *anxiety* (Cohen's $d = 0.40$), and *attitude towards mathematics* (Cohen's $d = 0.36$). Although his review finds these variables to be the "best" for comparison purposes, it is worth keeping in mind that the medium effect sizes in terms of Cohen's d (e.g., 0.40) are approximately transformed to Pearson product-moment correlations of around $r = 0.20$s (see Rosenthal, 1994). Another noteworthy limitation of Hattie's (2009) conclusion is the omission of self-efficacy from the list even though the author acknowledges the strong predictability of self-efficacy in relation to school outcomes. Several reports of the OECD (e.g., OECD, 2011, 2015) indicate a strong association between self-efficacy and student achievement. Other studies based on large-scale data (Lee, 2009; Lee & Stankov, 2013) and meta-analyses (Holden, Moncher, Schinke, & Barker, 1990; Multon, Brown, & Lent, 1991) have also shown that self-efficacy is potentially the best predictor of students' academic achievement.

1.1.2. Richardson et al. (2012)

Another major recent review of the literature has been conducted by Richardson et al. (2012). This meta-analysis of psychological correlates of academic performance among university students was based on 1105 independent correlations drawn from empirical studies published between 1997 and 2010. Fifty conceptually distinct constructs were analyzed in terms of university student grade point average (GPA), 42 of which belong to one of the following five research domains: *personality traits, motivational factors, self-regulatory learning strategies, students' approaches to learning*, and *psychosocial contextual influences*. Out of 50 measures reviewed, *performance self-efficacy*, which was defined as "the perception of academic performance capability" with an example item of "What is the highest GPA that you feel completely certain you can attain?" (p. 356), shows the largest correlation with GPA ($r = 0.59$). This correlation was higher than the correlations obtained between the university GPA and academic performance measures of: high school GPA ($r = 0.40$), intelligence ($r = 0.20$), and college entrance test performance in the Admission College Test ($r = 0.40$) and Standardized Admission Test ($r = 0.29$).

Other non-cognitive constructs that show moderately strong correlations with university GPA are: *academic self-efficacy* ($r = 0.31$, defined as one's "general perceptions of academic capability" with an example item of "I have a great deal of control over my academic performance in my courses", p. 356), *effort regulation* ($r = 0.32$, defined as making effort "when faced with challenging academic situations" with an example item of "I have enough self-discipline to complete", p. 357), and *grade goal* ($r = 0.35$, defined as "self-assigned minimal goal standards" with an example item of "What is the minimum [i.e., the least you would be satisfied with] percentage grade goal for the next test on a scale of 0% to 100%?", p. 357). However, most constructs included in their analyses show, at best, small correlations with GPA. According to Richardson et al. (2012), constructs that are typically studied in research domains of *learning strategies* and *motivation* theories did not reach a correlation of 0.20. Such constructs include: *attributional style* ($r = 0.01$), *intrinsic motivation* ($r = 0.17$), *extrinsic motivation* ($r = 0.01$), *mastery goal orientation* ($r = 0.10$), *performance goal orientation* ($r = 0.09$), *rehearsal/memorisation* ($r = 0.01$), and *deep approach to learning* ($r = 0.14$).

1.1.3. Stankov (2013)

The review by Stankov (2013) (see also Stankov & Lee, 2014, 2015) evaluated non-cognitive constructs spanning a broad range of educational (e.g., self-efficacy, motivation, learning strategies), cognitive (e.g., self-assessment), clinical (e.g., well-being and depression), and social (e.g., attitudes, values, social axioms and social norms) psychology. The predictability gradient hypothesis was introduced to suggest that non-cognitive processes can be ordered from those that show high correlations to those that have low or essentially zero correlations with academic and cognitive performance. The authors report that measures of *domain-specific self-concept* are often close to the borderline of being noteworthy, its correlations with student achievement typically being around mid $r = .20$s. More consistent and moderately strong correlations with academic achievement were found with *anxiety* (e.g., test anxiety or math anxiety, $r = .30$s) and *self-efficacy* (i.e., $r = .40$s). Measures of *item-based confidence*, which is typically assessed concurrently with some type of cognitive item performance (see Stankov, 2013, pp. 728–9 for the assessment procedure), are reported to correlate up to $r = 0.60$ with achievement.

1.1.4. Lee and Shute (2010)

Another recent comprehensive review of students' non-cognitive attributes was conducted by Lee and Shute (2010). Based on some 600 studies published between 1950 and 2010, the review provides a summary of empirical evidence of > 60 non-cognitive constructs. It concludes that only a dozen non-cognitive constructs appear to have direct, scientifically documented links to students' academic achievement at the K-12 school levels. The list includes *attitude, anxiety, self-beliefs, classroom behaviors, control strategies, elaboration strategies, engagement, enjoyment, extracurricular activity, homework, interest, metacognition, motivation, parental involvement, school climate, sense of belonging, self-confidence, self-concept, self-efficacy, student-teacher relationship, teacher efficacy, teacher support,* and *time spent on tasks*. These variables are further grouped into four major categories: student engagement, learning strategies, school climate, and social-familial influences, with the first two referring to personal dispositions and the latter two classified as social-contextual influences. Although the review successfully recognized a dozen non-cognitive constructs as particularly important for student learning, no specific claims were made about the superiority of any of the reviewed constructs as predictors of student achievement.

1.2. Aims of the present investigation

The general aim of the present study was to identify non-cognitive constructs that have direct and strong linear relationships to students' academic achievement. The four reviews described above had similar research aims, but our study has additional and broader objectives. First, we used the TIMSS and PISA non-cognitive assessment

[1] A large number of variables are included in the four reviews and in our current paper. Thus, providing definitions of each construct is not within the scope of this paper. Readers can obtain definitions of most constructs from the four review articles.

frameworks and constructs as the starting point for our review in order to include as many relevant constructs as possible and in so doing, to provide a comprehensive review of empirical findings based on the international datasets. Previous studies were largely reviews of published articles, the majority of which were not based on systematic, large-scale, international data. On the other hand, the non-cognitive constructs that were identified, advocated, and measured in the TIMSS and PISA were drawn from the selection processes agreed among the international and national research and assessment communities. The final selection of the constructs was seen as being scientifically important and practically relevant to students' educational outcomes in their own countries.[2] In this sense, large-scale international assessment databases such as TIMSS and PISA provide a unique opportunity to synthesize results that can be generalized to the > 60 systems and countries that participated in the assessments.

Second, we used five sets of data but analyzed each one separately, to maintain the consistency of various assessment settings of a single dataset as a whole (e.g., potential control over survey administration, assessment year, purposes). Although systematic analysis is performed in typical review papers, their data sources tend to be less systematic because they are drawn from individual studies, hence their comparability can be challenged. Thus, one of the main motivations of the present paper was to compare the findings of the previous major reviews with those from our study, which has a broader context not only in the nature of the sample (i.e., international) and actual sample coverage (i.e., large-scale) but also in the way the data were collected, that is, they were nationally representative samples of the same age/grade student group, most of which were surveyed within a few months of the actual assessment. Given that the self-belief constructs were identified as strong correlates of student achievement in all four reviews (Hattie, 2009; Lee & Shute, 2010; Richardson et al., 2012; Stankov, 2013), it was expected that similar results would emerge in the large-scale, international data. On the other hand, mixed results have been reported for the other non-cognitive constructs such as motivation, engagement, attitudes, school climate, social-familial influences, and learning strategies, so we had no particular expectations for these in the present study.

Third, we also aimed to contribute to the existing research literature by providing empirical evidence from the analysis of a vast number of non-cognitive constructs at a domain level. The constructs were grouped into research domains in a straightforward manner. For instance, a variable of "positive affect to math (TIMSS)" was placed into the domain of *affect*; "mathematics homework (TIMSS)" into *homework*; "valuing math (TIMSS)" into *value*; "mathematics self-efficacy (PISA)" into *self-beliefs*; and "disciplinary climate (PISA)" into *school climate*. Thirteen groups of educational psychology research domains appear in the TIMSS and PISA non-cognitive assessment: *affect, curriculum/content exposure, homework, learning and instructional time, motivational factors, personality traits, planned behavior, school climate, self-beliefs/social-cognitive theory, self-regulatory learning style/strategies, teacher behavior, value,* and *vocational interest* (see Table 1 for the construct-domain correspondence). Each domain can also be seen as a research or theoretical stream of educational psychology (e.g., motivation research).

Lastly, from a practical standpoint, we hoped that our analysis would provide evidence-based arguments to establish the relative strength of non-cognitive constructs for student achievement. Such information may ultimately be utilized by educational policy makers and the global assessment research community to design and implement educational initiatives to maximize student achievement and longer-term learning outcomes.

2. Methods

2.1. Data

The current study used the Student Background Questionnaire data from five cycles of TIMSS and PISA, i.e., TIMSS 2003, TIMSS 2007, TIMSS 2011, PISA 2003, and PISA 2012. Our analytic focus was on students' mathematics achievement and questions about mathematics learning.[3] We used the public-use data files, which were downloaded from the websites of http://timssandpirls.bc.edu/ (TIMSS) and https://www.oecd.org/pisa/ (PISA). Approval from the institutional review board (IRB) for research involving human subjects was not required by our institutions for secondary data analyses.

2.2. Participants

TIMSS selects participants based on grade level (i.e., fourth grade and eighth grade). The mean age of the TIMSS participants across the countries at the time of testing was 9.5 years for fourth-grade students and 13.5 years for eighth-grade students (Joncas & Foy, 2012). Only data for the eighth-grade students were analyzed because they are closer to the age of the students participating in PISA, allowing for comparisons between the findings from the two datasets. The total sample sizes of TIMSS participants were N = 237,833 from 49 countries in TIMSS 2003; N = 246,112 from 59 countries in TIMSS 2007; and N = 287,395 from 63 countries in TIMSS 2011. The eligibility requirements for a country to participate in TIMSS are a sample size of 4000 students, 150 schools for each target grade, and > 50% of the students from each classroom (Martin & Mullis, 2012).

PISA uses age-based sampling (i.e., 15 years old), designed to measure 15 years of accumulated domain-specific achievement regardless of school entry policies across different countries. Due to the age-based sampling method, variation in grade level exists across and within countries (i.e., from grades 7 to 13), but the majority of students in the PISA samples were in grade 9 or 10 at the time the survey was administered. The number of students participating in PISA has increased from N = 276,165 from 41 countries in PISA 2003 to N = 485,490 from 64 countries in PISA 2012. The minimum sample size requirement is about 4500 students and 150 schools per country, with each school meeting the minimum student sample size requirement of 20 (OECD, 2012).

We used the data from all TIMSS and PISA participating countries for a given assessment year. The total sample size across the five assessments was N = 1,532,995 students from around the world. The population of this study can be defined as students enrolled in a school system attending the 8th grade for TIMSS and, for PISA, students aged 15 years from the participating countries in each assessment.

2.3. TIMSS and PISA non-cognitive variables

Many of the questionnaire items were converted to scales[4]; the re-

[2] The construct selection and validation work in TIMSS and PISA includes a few years of comprehensive pilot and field testing, which involves the international committees and national representatives of all participant countries as well as the "best" researchers whose work on particular constructs is widely known. The final scale validation includes rigorous psychometric evaluation, which is conducted for all participant countries separately, and across all participating countries jointly. Final versions of the background questionnaires are reviewed by internationally renowned researchers and assessment specialists (e.g., see Turner & Adams, 2007, Fig. 1, p. 239 for key committee groups in the PISA project development).

[3] PISA was also conducted in 2006 and 2009. Non-cognitive constructs measured by the Student Background Questionnaire were about science and reading, respectively. Many comparable non-cognitive constructs were measured in the TIMSS in all cycles and the PISA assessments with mathematics-focus years, which were 2003 and 2012.

[4] The scales of both TIMSS and PISA were constructed by using the item characteristic estimates derived from the partial credit model of the item response theory (IRT) (see Martin, Mullis, Foy, & Arora, 2012 for TIMSS scale construction; see OECD, 2014, p. 314

Table 1
TIMSS and PISA non-cognitive constructs and major research domains in educational psychology.

Domain	Constructs and Measures	Representative studies
Affect	Positive affect to math (TIMSS), Positive affect to science (TIMSS)	Ackerman et al. (2010); Linnenbrink (2006); Wentzel, Russell, and Baker (2016)
Curriculum/content exposure	Math specific activity (TIMSS), Traditional classroom activity in math (TIMSS), Traditional classroom activity in science (TIMSS), Science experiment activity (TIMSS), Familiarity with mathematical concepts (PISA), Experience with pure mathematics tasks (PISA), Experience with applied mathematics tasks (PISA)	Alexander (1992); Alexander, Kulikowich, and Schulze (1994); Reynolds and Walberg (1992)
Homework	Mathematics homework (TIMSS), Science homework (TIMSS), Weekly time spent on math homework (TIMSS), Weekly time on science homework (TIMSS), Relative time spent on math homework (PISA)	Bempechat (2004); Cooper, Robinson, and Patall (2006)
Learning/instructional time	Time spent on math extra lessons (TIMSS), Time spent on science extra lessons (TIMSS), Total minutes of instructional time per week (PISA), Learning time in mathematics (PISA), Out-of-school study time (PISA), Learning time in reading (PISA), Ratio of math and total instructional time (PISA)	Lee and Shute (2010); Reynolds and Walberg (1992)
Motivation	Student engagement in math lessons (TIMSS), Student engagement in science lessons (TIMSS), Instrumental motivation for mathematics (PISA), Mathematics interest (PISA), External attributions to failure in mathematics (PISA)	Fredricks, Blumenfeld, and Paris (2004); Patrick, Ryan, and Kaplan (2007); Pintrich and DeGroot (1990); Wentzel (2015); Wigfield, Cambria, and Eccles (2012)
Personality	Openness for problem solving (PISA), Perseverance (PISA), Mathematics work ethic (PISA), Sense of belonging to school (PISA)	Duckworth et al. (2007); Duckworth and Yeager (2015); McGeown et al. (2016); Messick (1996); Poropat (2009)
Planned behavior	Attitude towards school (TIMSS), Attitude towards school: Learning outcomes (PISA), Attitude towards school: Learning activities (PISA), Mathematics behavior (PISA), Mathematics intentions (PISA), Subjective norms in mathematics (PISA)	Ajzen and Madden (1986); Elliott, Armitage, and Baughan (2003)
School climate	Feeling of school safety (TIMSS), Disciplinary climate (PISA)	Hoy and Hannum (1997); Lee and Shute (2010); Voelkl (2012); Wentzel and Watkins (2011)
Self-beliefs	Confidence with math (TIMSS), Confidence with science (TIMSS), Mathematics self-efficacy (PISA), Mathematics self-concept (PISA), Mathematics anxiety (PISA), Educational aspiration (TIMSS), Expected educational level of student (PISA)	Bandura (1997); Guay et al. (2003); Holden et al. (1990); Lee (2009); Lee and Stankov (2013); Marsh (1986); Schunk (1983); Stankov and Lee (2015); Zimmerman et al. (1992)
Learning strategies	Competitive learning (PISA), Co-operative learning (PISA), Use of control strategies (PISA), Use of elaboration strategies (PISA), Use of memorization strategies (PISA)	Eshel and Kohavi (2003); Hattie, Biggs, and Purdie (1996); Lee and Shute (2010); Lee and Stankov (2013); Pintrich and DeGroot (1990); Zimmerman (1990)
Teacher behavior	Math teacher's classroom management (PISA), Cognitive activation in mathematics lessons (PISA), Teacher student relations (PISA), Mathematics teacher's support (PISA), Teacher support (PISA), Teacher behavior: Teacher-directedness (PISA), Teacher behavior: Formative assessment (PISA), Teacher behavior: Student orientation (PISA)	Clotfelter, Ladd, and Vigdor (2006); Jussim, Robustelli, and Cain (2009); Lee and Shute (2010); Ware and Kitsantas (2007); Wentzel (2009)
Value	Valuing math (TIMSS), Valuing science (TIMSS)	Eccles and Wigfield (1995); Wigfield and Eccles (2000).
Vocational interest	Labor market information (PISA), Labor market information outside of school (PISA), Information about careers (PISA)	Lent, Brown, and Hackett (1994); Brown and Lent (2015)

Two constructs (i.e., parental involvement in TIMSS and grade repetition in PISA) were excluded because there were no other constructs to be classified into the same theoretical domain.

levant scale scores are available in the TIMSS[5] and PISA[6] public-use databases. We used the scale scores of the variables (in addition to the scale labels, as shown in Tables 2 and 3), that are available in the official databases of TIMSS[7] and PISA. A total of 65 non-cognitive variables was reviewed and analyzed in the present study: 22 non-cognitive constructs from the TIMSS and 43 from the

(*footnote continued*)
and Figs. 16.1 and 16.2 for more detailed descriptions of the IRT scale construction of the PISA scales). At the construct/variable level, there were 16 variables in TIMSS 2011, 19 in TIMSS 2007, and 18 in TIMSS 2003.

[5] The TIMSS scale items are readily available on the TIMSS website (http://timssandpirls.bc.edu/). In the public-use TIMSS databases, not all background questionnaire items were converted to scale scores. Our analyses of the TIMSS data were conducted with the scale-level scores. The only exceptions were four items, each of which was measured at the item level: the highest education level of parents, student self-expectation of educational attainment, time spent on extra lessons on mathematics, and time spent on extra lessons on science. The highest education level of parents was derived from a higher level of mother's or father's education and eight response categories were provided as: some primary or lower secondary education or did not go to school; lower secondary education; lower secondary education; post-secondary non-tertiary education; short-cycle tertiary education; Bachelor or equivalent; beyond Bachelor or equivalent; and I don't know. Student self-expectation of educational attainment was measured by one item with seven response categories of: finish lower secondary education; finish upper secondary education; finish post-secondary non-tertiary education; finish short-cycle tertiary education; finish Bachelor or equivalent; beyond Bachelor or equivalent; and I don't know. Time spent on extra lessons on mathematics and time spent on extra

(*footnote continued*)
lessons on science were each measured with four Likert-type categories of: never or almost never; sometimes; once or twice a week; and every or almost every day.

[6] The PISA public-use database contains scale scores for the majority of the background questionnaire variables. The PISA scales and the items constituting the scales are available on the PISA website (a section labelled PISA Products http://www.oecd.org/pisa/pisaproducts/). There were only two variables not measured by a scale: the highest education level of parents and student self-expectation of educational attainment. The highest education level of parents was derived from a higher level of mother's or father's education and each had eight response options: elementary school; middle or junior high school; high school equivalency or GED; high school diploma; vocational or technical certificate/diploma after high school; associate's degree; Bachelor's, master's, doctorate or professional degree such as law or medicine; and none of the above. Student self-expectation of educational attainment was measured by one item with five options to choose from: middle or junior high school; high school; vocational or technical certificate after high school; associate's degree; Bachelor's degree or higher.

[7] We used the scale labels, which were the same labels included in the official databases, publications and websites of TIMSS and PISA (see Chapter 3 in Martin et al., 2004 for the TIMSS 2003 scale definitions and descriptions, also available at https://timssandpirls.bc.edu/timss2003i/technicalD.html; see Chapter 3 in Olson et al., 2008

Table 2
Correlations between TIMSS Mathematics Achievement and Non-cognitive and SES variables (N = 287,395 for TIMSS 2011; N = 246,112 for TIMSS 2007; N = 237,833 for TIMSS 2003).

	Variable label in the TIMSS database	Variable description	Pan-cultural Level			Within-country Averages			Between-country Level		
			2011	2007	2003	2011	2007	2003	2011	2007	2003
1	BSBG05A-E	Home possessions	.462	.443	.459	.249	.275	.253	.776	.735	.774
2	BSDGEDUP	Parental education	.377	.353	.354	.313	.269	.287	.532	.582	.535
3	BSBG07	Educational aspiration	.270	.227	.299	.393	.368	.367	-.202	-.215	.073
4	BSDGSCM	Confidence with math	.253	.217	.208	.398	.397	.373	-.296	-.496	-.319
5	BSDGSBS	Feeling of school safety	.207	.177	.260	.106	.116	.113	.551	.484	.629
6	BSDGSLM	Positive affect to math	.065	.013	n/a	.258	.221	n/a	-.626	-.675	n/a
7	BSDMWKHW	Weekly time spent on math homework	.046	-.007	.013	.050	.048	.026	.008	-.152	-.082
8	BSDGSCS	Confidence with science	.039	.008	-.009	.208	.210	.190	-.621	-.688	-.631
9	BSDGSLS	Positive affect to science	-.017	-.075	n/a	.143	.087	n/a	-.765	-.696	n/a
10	BSDGSVM	Valuing math	-.038	-.025	-.085	.163	.122	.145	-.733	-.663	-.681
11	BSDGEML	Student engagement in math lessons	-.069	n/a	n/a	.145	n/a	n/a	-.785	n/a	n/a
12	BSBG12A-C	Attitude towards school	-.079	-.174	-.192	.08	.021	.033	-.712	-.721	-.736
13	BSDSWKHW	Weekly time on science homework	-.079	-.146	-.138	-.014	-.020	-.042	-.353	-.519	-.552
14	BSDGESL	Student engagement in science lessons	-.091	n/a	n/a	.104	n/a	n/a	-.842	n/a	n/a
15	BSBG11A-D	Parental involvement	-.113	n/a	n/a	-.008	n/a	n/a	-.552	n/a	n/a
16	BSDGSVS	Valuing science	-.116	-.136	-.178	.118	.095	.088	-.808	-.754	-.727
17	MATSPEC	Math specific activity	n/a	-.018	-.031	n/a	.032	.050	n/a	-.209	-.292
18	MATTRD	Traditional classroom activity in math	n/a	-.044	-.148	n/a	.098	.073	n/a	-.582	-.694
19	SCTTRD	Traditional classroom activity in science	n/a	-.06	-.226	n/a	.105	.003	n/a	-.662	-.710
20	MAHWK	Mathematics homework	n/a	-.094	n/a	n/a	-.012	n/a	n/a	-.255	n/a
21	SCEXPT	Science experiment activity	n/a	-.140	-.117	n/a	-.023	-.024	n/a	-.429	-.303
22	SCHWK	Science homework	n/a	-.145	-.187	n/a	-.037	-.007	n/a	-.417	-.509
23	BSBMEXTO	Time spent on math extra lessons	n/a	n/a	-.192	n/a	n/a	-.168	n/a	n/a	-.328
24	BSBSEXTO	Time spent on science extra lessons	n/a	n/a	-.259	n/a	n/a	-.178	n/a	n/a	-.462

The public data bases of TIMSS use slightly different variable labels across different years. The labels listed are based on TIMSS 2011. The variables/scales not measured in 2011 are based on the 2007 or 2003 variable labels. There were no systematic linear relationships between the scales' reliability and their predictive validity reported in this table. The Cronbach's Alpha values of the scales vary from country to country, with most of them showing the values between .60s and .80s. The psychometric properties and techniques behind the scale construction, including the scale reliability information, can be found in Chapters 3 and 13 of the TIMSS 2003 technical report (Martin, Mullis, & Chrostowski, 2004) and Chapter 12 of the TIMSS 2007 technical report (Olson, Martin, & Mullis, 2008), and in the section on creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales in the methods and procedures in TIMSS and PIRLS 2011 (Martin & Mullis, 2012), which is also available online: https://isc.bc.edu/methods/t-context-q-scales.html.

PISA.[8] In total, 2045 items were reviewed across the TIMSS and the PISA scales: 549 items in PISA 2012 and 319 items in PISA 2003, and 399 items in TIMSS 2011, 386 items in TIMSS 2007, and 392 items in TIMSS 2003. We reviewed the content of the scale items to ensure there was close alignment between the scale labels and commensurate items. Our own evaluation of the items indicates that the content of the items is represented in the scale labels (published by the TIMSS and PISA development teams).[9] The scale scores were coded to indicate that a

(footnote continued)
for the TIMSS 2007 scale definitions and descriptions, also available at https://timssandpirls.bc.edu/TIMSS2007/techreport.html; see the "Context Questionnaire Scales Detail" in Martin & Mullis, 2012 for the TIMSS 2011 scale definitions and descriptions, also available at https://timssandpirls.bc.edu/methods/t-context-q-scales.html; see Annex A1 in OECD, 2004, pp. 306–316 for the PISA 2003 scale definitions, also available at http://www.oecd-ilibrary.org/education/learning-for-tomorrow-s-world_9789264006416-en; and Annex A1 in OECD, 2013, pp. 192–207 for the PISA 2012 scale definitions, also available at https://doi.org/10.1787/9789264201170-en).

[8] The majority of the TIMSS non-cognitive variables remained the same across different years of assessment, which made the same 22 TIMSS non-cognitive variables available in all three years (2011, 2007, and 2003). There were 34 scale-level non-cognitive variables in PISA 2012 and 20 scale-level non-cognitive variables in PISA 2003. The 11 scale-level variables were measured at both times, resulting in 43 unique non-cognitive PISA variables that were analyzed in the present study.

[9] Our goal was to examine as many non-cognitive variables as possible and include even those that are remotely considered non-cognitive (e.g., student time spent on learning and learning strategies). On the other hand, excluded are those that cannot be considered non-cognitive even in the broadest sense, such as variables that are considered outside of students' control (e.g., birth month, birth year, gender, family structure, language use, immigration-related questions in both the TIMSS and the PISA). In addition, two groups of variables were excluded: computer use items in both the TIMSS and the PISA and leisure activities in the TIMSS, because the existing literature does not provide a

Table 3
Correlations between the PISA Mathematics Achievement and Non-cognitive and SES variables (N = 485,490 for PISA 2012; N = 276,165 for PISA 2003).

	Variable label in the PISA database	Variable description	Pan-cultural level		Within-country Averages		Between-country level	
			2012	2003	2012	2003	2012	2003
1	MATHEFF	Mathematics self-efficacy	.461	.458	.460	.480	.603	.455
2	HOMEPOS	Home possessions	.360	.486	.286	.352	.521	.763
3	FAMCON	Familiarity with mathematical concepts	.357	n/a	.404	n/a	.136	n/a
4	SISCED	Expected educational level of student	n/a	.312	n/a	.440	n/a	-.239
5	EXPUREM	Experience with pure mathematics tasks	.306	n/a	.299	n/a	.470	n/a
6	HISCED	Highest educational level of parents	.272	.332	.262	.251	.234	.492
7	SCMAT	Mathematics self-concept	.264	.275	.342	.345	-.404	-.264
8	DISCLIMA	Disciplinary climate	.209	.168	.193	.188	.420	.241
9	OPENPS	Openness for problem solving	.174	n/a	.270	n/a	-.524	n/a
10	TMINS	Total minutes of instructional time per week	n/a	.109	n/a	.102	n/a	.161
11	PERSEV	Perseverance	.108	n/a	.191	n/a	-.450	n/a
12	MATINTFC	Mathematics intentions	.084	n/a	.123	n/a	-.133	n/a
13	MMINS	Learning time in mathematics	.070	.012	.109	.054	.037	-.095
14	INFOJOB1	Labor market information	.061	n/a	.122	n/a	-.243	n/a
15	EXAPPLM	Experience with applied mathematics tasks	.054	n/a	.074	n/a	.062	n/a
16	ATSCHL	Attitude towards school: Learning outcomes	.052	-.053	.103	.047	-.418	-.767
17	CLSMAN	Math teacher's classroom management	.046	n/a	.092	n/a	-.297	n/a
18	BELONG	Sense of belonging to school	.038	.032	.074	.056	-.235	-.085
19	INSTMOT	Instrumental motivation for mathematics	.030	.002	.133	.120	-.524	-.602
20	OUTHOURS	Out-of-school study time	.023	n/a	.051	n/a	-.151	n/a
21	LMINS	Learning time in reading	.019	n/a	.013	n/a	.048	n/a
22	MATWKETH	Mathematics work ethic	.019	n/a	.112	n/a	-.574	n/a
23	ATTLNACT	Attitude towards school: Learning activities	.013	n/a	.076	n/a	-.500	n/a
24	INFOJOB2	Labor market information outside of school	.011	n/a	.064	n/a	-.545	n/a
25	INTMAT	Mathematics interest	.009	-.003	.147	.161	-.535	-.775
26	COGACT	Cognitive activation in mathematics lessons	-.024	n/a	.041	n/a	-.472	n/a
27	INFOCAR	Information about careers	-.034	n/a	.013	n/a	-.438	n/a
28	STUDREL	Teacher student relations	-.051	-.087	.009	.004	-.467	-.612
29	MTSUP	Mathematics teacher's support	-.061	n/a	.008	n/a	-.478	n/a
30	TEACHSUP	Teacher support	-.073	-.091	.014	-.015	-.475	-.613
31	MATBEH	Mathematics behavior	-.078	n/a	.065	n/a	-.546	n/a
32	SUBNORM	Subjective norms in mathematics	-.121	n/a	-.028	n/a	-.548	n/a
33	TCHBEHTD	Teacher behavior: Teacher-directedness	-.121	n/a	-.051	n/a	-.532	n/a
34	FAILMAT	External attributions to failure in mathematics	-.160	n/a	-.149	n/a	-.365	n/a
35	TCHBEHFA	Teacher behavior: Formative assessment	-.169	n/a	-.113	n/a	-.565	n/a
36	REPEAT	Grade repetition	-.294	n/a	-.267	n/a	-.231	n/a
37	TCHBEHSO	Teacher behavior: Student orientation	-.330	n/a	-.236	n/a	-.708	n/a
38	ANXMAT	Mathematics anxiety	-.365	-.378	-.342	-.332	-.542	-.631
39	COMPLRN	Competitive learning	n/a	-.075	n/a	.063	n/a	-.667
40	COOPLRN	Co-operative learning	n/a	-.084	n/a	-.012	n/a	-.614
41	CSTRAT	Use of control strategies	n/a	-.047	n/a	.057	n/a	-.656
42	ELAB	Use of elaboration strategies	n/a	-.135	n/a	.028	n/a	-.826
43	MEMOR	Use of memorization strategies	n/a	-.104	n/a	-.020	n/a	-.748
44	PCMATH	Ratio of math and total instructional time	n/a	-.114	n/a	-.082	n/a	-.186
45	RMHMWK	Relative time spent on math homework	n/a	-.224	n/a	-.196	n/a	-.440

The variable names and labels were unchanged from those in the PISA's public-use data files. "n/a" means that the data is not available in the particular assessment year. There were no systematic linear relationships between the scales' reliability and their predictive validity reported in this table. The Cronbach's Alpha values of the scales vary from country to country with the OECD averages ranging from upper .60s to lower .90s and most of them being in the range between .70s and .80s. The reliability estimates can be found in Chapter 17 of the PISA 2003 Technical Reports (OECD, 2005) for the PISA 2003 scales (available at: https://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/35188570.pdf) and Chapter 16 of the PISA 2012 Technical Reports (OECD, 2014) for the PISA 2012 scales (available at: https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf).

student's higher score meant a higher level of the construct being measured.

2.4. Academic achievement measures in TIMSS and PISA

We correlated mathematics scores with students' non-cognitive scale scores. The cognitive domain tests in TIMSS and PISA tend to show high correlations across the tested school subjects. For instance, in PISA 2012, the inter-subject correlations were $r = 0.86$ between reading and mathematics, $r = 0.90$ between mathematics and science, and $r = 0.88$ between reading and science (OECD, 2014, p. 230). Similarly, the corresponding correlations in the PISA 2003 were $r = 0.77$ between reading and mathematics, $r = 0.82$ between mathematics and science, and $r = 0.83$ between reading and science (OECD, 2005, p.189). Thus, it is fair to say that the mathematics scores can be used as a proxy for students' overall academic achievement. In both TIMSS and PISA assessments, students' responses on the achievement tests were estimated with plausible values, which were then converted into scale scores. The mathematics scores are scaled to have a mean of 500 and a standard deviation of 100 in both TIMSS and PISA.[10]

2.5. Socio-economic status (SES) measures as benchmark variables

It is generally accepted that students' family SES measures are moderately strong correlates of academic achievement (e.g., Sirin, 2005). Thus, the correlations between the SES measures and the PISA and TIMSS achievement scores are used in the present study as a benchmark against which to judge predictive validity of non-cognitive variables. To maintain consistency between the TIMSS and the PISA data analyses, we employed two SES-related variables that were measured in both assessments—parental education and home possessions.[11] Many empirical studies of the relationships between SES and student achievement were conducted with students from culturally similar groups (e.g., Anglo and/or living in Western countries), so the international data of TIMSS and PISA provides a unique opportunity to assess the effects of SES on student achievement across many diverse cultural and national contexts.

(*footnote continued*)
clear direction pointing to the direct, positive relationship between computer use/leisure activities and student achievement (e.g., Hunley et al., 2005 on computer use; Kleiber, Larson, & Csikszentmihalyi, 1986 on leisure). Overall, the variable selection process led us to exclude only a handful of variables in the entire TIMSS and PISA databases.

[10] Scaling the TIMSS mathematics tests scores were originally established with the TIMSS 1995 data by setting the mean of all participant countries at 500 and the standard deviation at 100. To examine the trends in data across years, concurrent calibration was used for the TIMSS assessments in subsequent years (see Foy, Brossman, & Galia, 2012, p. 4 for more detailed descriptions of the scale setting methodology of the TIMSS achievement scores). Similarly, to compare the PISA 2003 and PISA 2012 scores, the scaling of the PISA 2012 mathematics achievement scores was conducted based on the PISA 2003 mathematics scores. The scale was established with the PISA 2003 mathematics mean at 500 and the standard deviation at 100 for the 30 OECD countries participating in the 2003 assessment (see OECD, 2012, p.159 for more detailed descriptions of the scale setting methodology of the PISA achievement scores). In this study, we reported the results based on the analysis of the first plausible values of the overall mathematics achievement as the mathematics scores for the TIMSS and PISA data. We have checked our results with the results obtained using the large-scale methodology, i.e., using student weight and replicates with the five plausible values for mathematics scores (i.e., the jackknife repeated replication (JRR) method for TIMSS and the balanced repeated replication (BRR) method for PISA). The differences are minor, i.e., within decimal points.

[11] Three SES variables were measured in the TIMSS assessments: parental education, home possessions, and amount of books at home. The home possessions scale contained the item about number of books, and thus the home possessions scale (which is broader than just books), along with the parental education levels were used in the present study. The PISA database, on the other hand, contained 10 SES scales, and among them two SES variables—parental education and home possession—were measured in both TIMSS and PISA and thus employed in our study.

2.6. Analytic strategies

There were three distinct stages of analysis for the present investigation: (a) selection of non-cognitive variables for inclusion in each step, (b) calculation of bivariate correlations between students' non-cognitive variables and mathematics achievement scores, and (c) two-level Hierarchical linear modelling (HLM) to arrive at a parsimonious set of non-cognitive predictors (at Level 1 and Level 2) to explain student academic performance at the individual and group (country) levels.

2.6.1. Correlations: noteworthy versus unimportant

Given that the number of variables examined in this study was quite substantial, we report only the summary results from our analysis. We first report the raw bivariate correlations, which are easily understood and can be interpreted as indicating direct effect sizes. Given the large sample sizes of the present study, we paid attention to the actual effect sizes (i.e., correlation and standardized regression coefficients) and not to the conventional statistical significance criteria relying on a 0.05 or 0.01 cut-off point (because very low coefficients are likely to be statistically significant). We relied on Cohen's (1988) guidelines for practically meaningful effect sizes in social science research. < 5% of common variance between two variables is considered to be of no practical importance. This 5% shared variance corresponds approximately to a correlation of $r = 0.2$. We used this criterion to draw conclusions about our findings.

2.6.2. Correlations: pan-cultural, average within-countries and between-countries

For each non-cognitive variable, three sets of correlation coefficients are reported: (a) pan-cultural (i.e., calculated based on individual students irrespective of their country of origin), (b) within-country correlation averages (i.e., calculating the averages of the within-country correlations that are obtained in each country), and (c) between-country correlations (i.e., based on the country aggregate scores to obtain the correlations with a country as a unit of analysis). Research adopting both pan-cultural and within-countries correlation averages tended to demonstrate similar results (see Stankov, 2011). Previous empirical studies also show that the between-countries correlations tend to be substantially higher than pan-cultural or within-countries correlations (e.g., Stankov & Lee, 2016). The three sets of correlations should provide stronger and more meaningful evidence for which non-cognitive constructs are important for individual students and students as a group (i.e., within a country). It would allow for the examination of any discrepancies in the correlations at the individual-versus-group level. The information garnered from the between-countries correlations also provides insight into selection and interpretations of the HLM findings at the group (country) level.

2.6.3. Hierarchical linear modelling (HLM)

Hierarchical linear models (HLM) were created and tested for each of the five datasets to identify the best non-cognitive predictors of mathematics achievement. The HLM analyses take into account the fact that students' responses had a nested structure (i.e., some students belong to the same country). HLM also allows for the examination of relative predictability of the non-cognitive measures simultaneously at the within-country individual level (Level 1) and between-country level (Level 2). Specific considerations in building a series of HLM models include: (a) all variables at both Level 1 and Level 2 are statistically significant; (b) deviance tests should show that a subsequent, more complex model is significantly different from the previous nested model; and (c) percentages of variance explained in the dependent variable are > 10%. Full information maximum likelihood procedure was used to estimate the HLM parameters. The final estimations of the fixed effects were analyzed with robust standard errors. Grand mean centering was used for all variables entered as predictors in the HLM because no a priori assumption was made about the individuals' relative group standings on the dependent variable (see Snijders & Bosker, 2011 on centering). Of the many models tested, only the final HLM model results are presented in this paper. The HLM analyses were carried out using the HLM version 7.0 (Raudenbush et al., 2011).

3. Results

This section presents the results from: correlations between mathematics and the TIMSS and PISA measures and the HLM analysis of the TIMSS and PISA data. Three types of correlations—pan-cultural, within-country averages, and between-country level—are organised in one table for each assessment (i.e., Table 2 for TIMSS; Table 3 for PISA). The variables are presented in descending order from those showing the largest positive to those showing the smallest and/or negative pan-cultural correlations with the mathematics achievement scores based on the findings for the most recent assessment years, i.e., TIMSS 2011 and PISA 2012. Unless otherwise specified, the results are illustrated with data from these two most recent assessments. The horizontal dotted lines indicate the variables showing pan-cultural level correlations > 0.20. Our comments will focus on these variables. The two SES variables *home possessions* and *parental education* are included primarily to compare their effect sizes to the effect sizes of the non-cognitive variables themselves (i.e., the main interest of this study).

3.1. Pan-cultural correlations in the TIMSS data

The first three columns of Table 2 present pan-cultural correlations between the non-cognitive and SES variables and mathematics achievement in the TIMSS data. Two SES variables—*home possessions* ($r = 0.462$) and *parental education* ($r = 0.377$)—showed the highest correlations, followed by student self-expected level of *educational aspiration* ($r = 0.270$), *confidence in learning mathematics*[12] ($r = 0.253$), and student feeling of *school safety* ($r = 0.207$). Among all 24 variables, only the top five had correlations > 0.20.

On the other hand, the majority of the TIMSS non-cognitive variables did not surpass $r = +0.10$ correlation and about one-third of them showed seemingly paradoxical findings (very low, near zero, and/or negative correlations with student achievement, see Table 2). This is a surprising result given the context of questionnaire development of the large-scale international assessments, in which various national and international expert groups and committees from > 60 countries contributed to the variable selection process. Overall, the correlation patterns are remarkably consistent across 2011, 2007 and 2003 (e.g., *home possessions*: *rs* = 0.462, 0.443, and 0.459 in 2011, 2007, and 2003, respectively).

3.2. Averages of the within-country correlations in the TIMSS data

The middle three columns of Table 2 present averages of the within-country correlations. As was the case with the pan-cultural correlations, there are close similarities in the findings across the three years. There were also close similarities between the pan-cultural and within-country analyses results in terms of which variables showed moderately strong correlations with achievement. Only four variables had moderately strong correlations in both analyses ($r \geq 0.20$): *home possessions* and *parental education* (two SES variables) and *students' educational aspiration*, and *confidence in mathematics* (two self-beliefs variables).

There were also some minor but noteworthy differences in the two sets of correlation results. The most notable difference is in the changes of the rank-orders of the first five variables (i.e., those with the correlations > 0.20 in the pan-cultural analysis). *Home possessions*, which ranked first at the pan-cultural level ($r = 0.462$), shows a rather modest average within-country correlation ($r = 0.249$). *Confidence in mathematics* ($r = 0.398$) and student *educational aspiration* ($r = 0.393$), which slightly surpassed the SES-related variable correlations of *parental education* ($r = 0.313$) and *home possessions* ($r = 0.249$), are now the strongest correlates of mathematics in the within-country analysis. These patterns are consistent in the TIMSS 2007 and 2003 data, as well.

Smaller effect sizes of SES variables (*home possessions* and *parental education*) on achievement in the within-country average correlation suggest that SES of a country (e.g., wealthier or developing) may have contributed to the diminished effects of the SES variable on achievement. For instance, students who live in a developed country are likely to derive advantages from the development/wealth of their country (such as school building conditions, infrastructure to commute to school, etc.). Thus, the effect of SES on achievement within the particular country may have been weakened to some degree. The same logic can also be applied to students in developing countries where the effect of SES on achievement is potentially weakened by the general living conditions in the countries. Thus, weaker correlations can be shown when the SES-achievement relationship is examined within a country.

On the other hand, *confidence in mathematics* and students' *educational aspiration* (two self-belief variables) showed the opposite pattern: stronger correlations in the within-country averages ($r = 0.398$ and $r = 0.393$, respectively) than their corresponding pan-cultural level correlations ($r = 0.253$ and $r = 0.270$, respectively). Among students from the same country, their non-cognitive dispositions seem to matter more in attaining higher achievement. Thus, it appears that students' self-beliefs and achievement are better aligned with each other when the two variables are examined in the within-country context.

3.3. Between-country correlations in the TIMSS data

Correlations based on the country-aggregate scores (between-countries correlations) are presented in the last three columns of Table 2. Again, the correlational patterns are similar across the 2011, 2007, and 2003 TIMSS data. The between-country correlations of most variables are amplified substantially when they are compared to the corresponding pan-cultural or within-country correlations. It is a well-known phenomenon that correlations based on aggregate scores tend to be higher than correlations of the original (non-aggregated) scores although they should, in theory, produce the same/similar sizes and directions of correlations (see Ostroff, 1993).

Among the moderately strong correlates of student achievement at the pan-cultural and within-country levels (i.e., those listed from #1 to #5), a couple of noteworthy observations can be made. Two self-beliefs variables, *educational aspiration* and *confidence in mathematics,* appear not to have the inflated correlations with achievement scores. However, the direction of their correlation is now reversed from positive at the pan-cultural and within-country levels to negative at the between-country level. This means that a country reporting a greater level of confidence actually had a lower score in mathematics.

Across all 24 variables only three, *home possessions* ($r = 0.776$), *parental education* ($r = 0.532$), and *students' feeling of school safety* ($r = 0.551$), showed positive associations with mathematics at the between-country level. The remaining non-cognitive variables in the TIMSS had negative (or near zero) correlations with mathematics achievement at the between-country level. In fact, different signs of correlations at the individual- and group-level of analysis present seemingly contradictory results. This phenomenon has long been noted and different labels have been used to explain it, such as "attitude–achievement paradox" (Mickelson, 1990), "paradoxical relationship" (Van de Gaer, Grisay, Schulz, & Gebhardt, 2012), or "anomaly" (Kyllonen & Bertling, 2013). While several attempts were made in PISA

[12] The scale labelled as BSDGSCM in the TIMSS, "confidence in learning mathematics", is similar to the PISA mathematics self-concept scale (SCMAT). An example of the TIMSS' self-concept items is "*I learn things quickly in mathematics*". An example of the PISA's self-concept scale (SCMAT) is "*I learn mathematics quickly*". We retained the label of confidence for the TIMSS's scale. The items used in the PISA scale of mathematics self-efficacy (MATHEFF) are unique to the PISA and no similar items/scales were used in the TIMSS.

2012 to resolve this paradox (e.g., use of anchoring vignettes; see Kyllonen & Bertling, 2013), the interpretations of the results based on the proposed "new" methods are still open to question (Stankov, Lee, & von Davier, 2017). Currently, no clear-cut interpretations seem to be available to suggest whether the observed differences in the signs of correlations are due to "true" cross-cultural differences operating at the higher level, or statistical/methodological artefacts and response bias at the group/country level. For the time being, what can be concluded is that the variables not displaying the "paradoxical relationship" (i.e., same directionality, whether positive or negative) may be considered as more consistent and, thus, arguably more robust as correlates of achievement. In the example of SES variables such as *home possessions*, it can be stated that students and countries with higher SES would show higher achievement because the positive effects of SES on achievement are consistent at the within- and between-country levels.

3.4. Pan-cultural correlations in the PISA data

The first two columns of Table 3 present the pan-cultural correlations between the PISA mathematics scores and non-cognitive and SES-related variables based on the 2012 and 2003 data. There was considerable similarity across the results of these two years, not just in the strength and direction, but also in the actual size of the correlations obtained with the same variables (e.g., $r = 0.461$ in 2012 and $r = 0.458$ in 2003 for *mathematics self-efficacy*). As was the case with the TIMSS, only a handful of variables showed moderately strong and positive correlations (i.e., $r \geq +0.20$) with mathematics achievement: two SES-related variables—*home possessions* ($r = 0.360$) and *parental education* ($r = 0.272$); three self-belief variables—*mathematics self-efficacy* ($r = 0.461$), *expected educational level of student* ($r = 0.312$ in 2003), and *mathematics self-concept* ($r = 0.264$); two variables relating to curriculum/content exposure of students—*familiarity with mathematics concepts* ($r = 0.357$) and *experiences with pure mathematics tasks* ($r = 0.306$); *disciplinary climate at school* ($r = 0.209$); and *mathematics anxiety* ($r = -0.365$) and *grade repetition* ($r = -0.294$). The last two variables were expected to be negatively correlated with mathematics, and the correlations were in the expected direction. Overall, three variables stand out: *mathematics self-efficacy* ($r = 0.461$) and *mathematics anxiety* ($r = -0.365$) as the best correlates of mathematics achievement, followed by *home possessions* ($r = 0.360$) in the 2012 data.

3.5. Averages of the within-country correlations in the PISA data

Averages of the within-country correlations between the PISA variables and mathematics achievement scores are presented in the middle two columns of Table 3. The overall pattern of the within-country correlations did not differ to any great extent from the correlational results of the pan-cultural analysis. As in the pan-cultural analysis, the strongest correlates of mathematics achievement were *mathematics self-efficacy* ($r = 0.460$) and *expected educational level of student* ($r = 0.440$ in 2003). Two SES variables, *home possessions* and *parental education*, had correlations in the 0.20 range: $r = 0.286$ and $r = 0.262$, respectively.

There appear to be three groups of variables when the within-country averages are compared to the corresponding pan-cultural correlations. The first group belongs to the variables that produced highly close correlations in both analyses; namely, *mathematics self-efficacy* ($r = 0.461$ in the pan-cultural analysis and $r = 0.460$ as the average of the within-country estimates) and *experiences with pure mathematics tasks* ($r = 0.306$ and $r = 0.299$). Almost identical size of correlations obtained from the within-country and pan-cultural analyses suggest that the effects of those variables on student achievement are largely unaffected by country membership and can be treated as "universal". The same pattern was observed in both the 2012 and 2003 PISA data as well (see Table 3).

The second group includes the variables for which the correlations are weaker in the within-country analysis compared to the pan-cultural analysis. The SES-related variables belong to this group, as indicated in the correlations of *home possessions* (from $r = 0.360$ to $r = 0.286$ in 2012 and $r = 0.486$ to $r = 0.352$ in 2003). A similar pattern was observed with *parental education* (from $r = 0.272$ to $r = 0.262$ in 2012 and $r = 0.332$ to $r = 0.251$ in 2003). The same pattern of results was also obtained with the SES-variables in the TIMSS 2011, 2007, and 2003 data, as presented earlier.

The third group consists of the variables showing stronger within-country correlations compared to the pan-cultural correlations. Of the eight variables that have PISA 2012 pan-cultural correlations higher than 0.20 in Table 3, three belong to this group: *familiarity with math concept* (from $r = 0.357$ to $r = 0.404$), *expected educational level of student* (from $r = 0.312$ to $r = 0.440$ in 2003), and *mathematics self-concept* (from $r = 0.264$ to $r = 0.342$). This pattern of correlations may indicate that there is closer alignment between student achievement and perceptions of their own abilities/expectations when the relationships are examined within a country as opposed to across different countries. Similar patterns were also observed in the TIMSS 2011, 2007, and 2003 assessments with respect to *confidence with mathematics* and *students' educational aspiration*.

3.6. Between-country correlations in the PISA data

Only eight out of 46 variables demonstrated the same direction of relationship with achievement at the between-level of analysis as was obtained in the pan-cultural/within-country analyses. They are: *home possessions* and *parental education* (two SES-related variables), *self-efficacy in mathematics* and *disciplinary climate* (two non-cognitive variables included in both the PISA 2012 and 2003), *familiarity with mathematical concepts* and *experience with pure mathematics tasks* (two non-cognitive variables newly introduced in PISA 2012), and *grade repetition* and *mathematics anxiety* (two variables showing a negative relationship with mathematics). For instance, *parental education levels* and *grade repetition* showed fairly consistent results in the PISA 2012 data across the pan-cultural, within-country, and between-country analyses ($r = 0.272$, $r = 0.262$, $r = 0.234$ and $r = -0.294$, $r = -0.267$, $r = -0.231$, respectively). On the other hand, the direction and size of the between-country correlations of the majority of variables differ drastically from their corresponding pan-cultural or within-country correlations (i.e., substantially magnified negative correlations), as was observed in the TIMSS data.

3.7. HLM analysis results

The HLM results pertaining to TIMSS 2011 data are presented first and in length, while findings of the 2007 and 2003 data are briefly summarized. This is because the three datasets produced highly consistent results and presenting the HLM findings of the three datasets would be redundant. Similarly, the PISA 2012 results are more fully presented than the PISA 2003 results because the findings in the two datasets are similar. Standardized coefficients and corresponding *t*-ratios are presented as fixed effects—rather than unstandardized coefficients—to facilitate comparison across predictive values of the variables included in the model.

3.7.1. Evidence from TIMSS 2011 data

HLM variable selection was based on the review of all variables measured in the TIMSS databases as well as their results in the correlational analysis (Table 2). This process led to the conclusion that the majority of variables will not be strong predictors of mathematics because only five variables passed the threshold of showing pan-cultural correlations greater than $r = 0.20$. They are: *home possessions*, *parental education*, *students' educational aspiration*, *confidence in mathematics* and *students' feeling of school safety* (see Table 2). These variables were

Table 4
Hierarchical linear regression modelling of TIMSS 2011, 2007, and 2003 data.

	TIMSS 2011 (N = *189,520*)		TIMSS 2007 (N = *146,194*)		TIMSS 2003 (N = *140,224*)	
Fixed Effects (Level 1)	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio
Intercept, γ_{00}	474.74	95.43	465.01	94.96	479.70	83.55
Home possessions, γ_{10}	0.12	8.91	0.13	−11.82	0.11	−8.13
Parent education, γ_{20}	0.13	15.35	0.10	12.13	0.11	−10.48
Educational Aspiration, γ_{30}	0.20	20.52	0.18	16.61	0.17	16.97
Confidence in math, γ_{40}	0.24	28.78	0.24	−26.65	0.21	−21.94
School safety, γ_{50}	0.05	6.02	0.05	−8.89	0.06	7.32
Fixed Effects (Level 2)						
Home possessions, γ_{01}	0.59	8.96	0.45	7.47	0.60	7.88
Confidence in math, γ_{02}	−0.50	−6.27	−0.59	−7.28	−0.36	−5.92
Random Effects (Variance Components)						
Level 1, r_{ij}	5205.98		4885.43		4480.81	
Level 2, u_0	1194.76		1201.39		1560.54	

The differences in the sample size between correlational and HLM analyses are due to missing values at Level 2 because countries with missing data at Level 2 are not allowed in HLM. Standardized coefficients are presented for fixed effects. All variables are grand mean centered. All the coefficients and their associated *t*-ratios for the fixed effects are significant at the $p < .001$ level. All Level 2 variance components are significant at the $p < .001$ level.

considered and selected for the HLM analysis as potentially strong predictors of mathematics achievement. All five variables were entered at individual-level (Level 1). They were all entered at Level 2 as well, but variable selection for the final model was made such that all variables at both levels are not only significant but also show moderately strong effect sizes.[13] The backward elimination process was used for selection of Level 2 predictors of the HLM analysis to explain the country mean differences. The results of the HLM final model are presented in Table 4.

The standardized coefficients show relative effect sizes for mathematics beyond the contributions of the other variables in the model. Each of the five variables at Level 1 was a significant individual-level (within countries) correlate of mathematics ($p < .001$) even when they were simultaneously considered. The fixed effect of *confidence in mathematics* was the largest (0.24, $p < .001$), followed by *students' educational aspiration* (0.20, $p < .001$). Thus, an increase of 0.24 standard deviations (about 24 points) in the TIMSS 2011 mathematics scores was associated with one unit of standard deviation increase in *confidence in mathematics* when the other effects were held constant. Similarly, an increase of 0.20 standard deviations (about 20 points) in the TIMSS 2011 mathematics score was associated with one unit of standard deviation increase in *students' educational aspiration* when the other effects were held constant. The fixed effects of two SES variables, *home possessions* (0.12, $p < .001$) and *parental education* (0.13, $p < .001$), were about half the size of the fixed effects of the two non-cognitive variables. A point to remember is that, in the bivariate correlations (Table 2), two SES-related variables showed stronger correlations with mathematics than *students' educational aspiration* or *confidence in mathematics*. In the HLM, however, the effects of *confidence* and *aspiration* were stronger than the effects estimated for *home possessions* or *parental education* on mathematics achievement, suggesting that two measures of self-belief are better predictors of achievement than measures of SES. The effect of *feeling of school safety* was comparatively smaller (0.05, $p < .001$).

The effect sizes of the two Level 2 variables were similar in magnitude but opposite in their directions: *home possessions* (0.59, $p < .001$) and *confidence in mathematics* (−0.50, $p < .001$). There was an increase of 0.59 (about 59 points) standard deviations in mathematics performance at the country level per one-unit increase of standard deviations of *home possessions*. In addition, there was a decrease of 0.50 (about 50 points) standard deviations in mathematics performance at the country level per one-unit increase of standard deviations of *confidence in mathematics*. Overall, the HLM model explained 30% of the variance at Level 1 (individuals within countries), 72% of the variance at Level 2 (between countries), and 45% of the total variance in mathematics in TIMSS 2011.

3.7.2. Evidence from TIMSS 2007 and 2003

The same approach to variable selection and model construction as used for TIMSS 2011 was adopted for the 2007 and 2003 TIMSS data. Highly consistent results were obtained in the HLM analysis across the three years. As can be seen in Table 4, the close resemblance in the HLM findings across the three years was observed in not only the relative strength of the predictors, but also in the actual magnitude of the fixed and random effect sizes. The strongest Level 1 predictor was *confidence in mathematics*, with fixed effect sizes of 0.24 in 2011, 0.24 in 2007, and 0.21 in 2003. The next strongest effect was found in *students' educational aspiration*: 0.20 in 2011, 0.18 in 2007, and 0.17 in 2003. The two SES-related variables, *home possessions* and *parental education* showed about half the fixed effect sizes of *confidence in mathematics* and *students' educational aspiration* at Level 1, which was observed across all three years. As for Level 2 predictors, two variables, *home possessions* and *confidence in mathematics*, were the best predictors of the country-level mean mathematics scores in both 2007 and 2003, as was the case in the 2011 data.

3.7.3. Evidence from PISA 2012

The approach used in the TIMSS HLM analysis was also applied to create and test hierarchical linear models of the PISA data. The variable selection for the HLM analysis started with a review of the correlational results presented in Table 3, using the selection criterion of correlations greater than $r = 0.20$ at the pan-cultural level in either 2012 or 2003. This process led to the selection of nine variables in the PISA 2012 data and seven variables in the PISA 2003 data. Among them, *mathematics self-efficacy*, *home possessions*, *parental education*, *mathematics self-concept*, *disciplinary climate at school*, and *mathematics anxiety* were measured in both PISA 2003 and 2012, and thus included in the modelling. In addition, the students' own *self-expected educational level* from the

[13] The consideration for Level 2 variables to predict the country-level means was based on the following process. Among the five variables, *home possessions* showed the largest and positive correlation with mathematics at the between-country level (i.e., $r = 0.776$ in 2011 shown in Table 2). Thus, it was firstly chosen as a Level 2 predictor. The remaining four variables were also evaluated as Level 2 predictors. Now, there are 10 variables simultaneously to consider across both levels (i.e., five Level 2 variables to predict the country mean scores and five Level 1 variables to predict individual-level mathematics scores within countries). Thus, the backward elimination method was employed by entering all 10 variables together and then selecting only those that are significant at both Level 1 and Level 2. Only two predictors at Level 2, *home possessions* and *confidence in mathematics*, were significant at $p < .001$ while all the Level 1 predictors remained significant at $p < .001$.

Table 5
Hierarchical linear regression modelling of PISA 2012 and 2003 data.

	PISA 2012 (N = *146,650*)		PISA 2003 (N = *258,135*)	
Fixed Effects (Level 1)	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio
Intercept, γ_{00}	476.37	109.13	493.24	114.99
Home possessions, γ_{10}	0.11	10.98	0.13	13.62
Parent education, γ_{20}	0.11	10.43	0.05	6.32
Grade repeat, γ_{30}	−0.19	−18.48	–	–
Student expectation, γ_{30}	–	–	0.22	18.54
Math efficacy, γ_{40}	0.25	18.02	0.25	21.36
Math anxiety, γ_{50}	−0.17	−16.11	−0.12	−11.20
Math self-concept, γ_{60}	0.05	4.88	0.04	3.83
School climate,, γ_{70}	0.07	11.75	0.07	16.57
Fixed Effects (Level 2)				
Home possessions, γ_{01}	0.28	4.24	0.57	6.43
Math self-concept, γ_{02}	−0.49	−5.72	−0.34	−3.91
Random Effects (Variance Components)				
Level 1, r_{ij}	4855.61		5007.49	
Level 2, u_0	1197.64		746.12	

Grade repeat (REPEAT) was measured in PISA 2012 only. *Expected educational level of student* (SISCED) was measured in PISA 2003 only. Due to the rotation design employed in PISA 2012, only about one-third of the sample provided responses to seven scales employed as predictors in HLM analysis (see OECD, 2014 for rotation design). Standardized coefficients are presented as the fixed effects. All variables are grand mean centered. All the coefficients and their associated *t*-ratios for the fixed effects are significant at the $p < .001$ level. Level 2 variance components are significant at the $p < .001$ level.

2003 data and *grade repetition* from the 2012 data were included while two variables that were newly added in 2012, *familiarity with mathematics concepts* and *experiences with pure mathematics tasks,* were excluded in the HLM analysis because no similar constructs were present in the 2003 data.

Table 5 presents the final models of HLM for the PISA data. In the 2012 data, seven variables, *mathematics self-efficacy*, *home possessions*, *parental education*, *grade repetition*, *mathematics self-concept*, *mathematics anxiety* and *disciplinary climate at school*, were entered as individual-level (Level 1) predictors of the PISA 2012 mathematics and were all statistically significant at the $p < .001$ level. *Mathematics self-efficacy* showed the largest standardized coefficient (0.25, $p < .001$), indicating that one unit of standard deviation increase in student perception of *mathematics self-efficacy* was associated with an increase of 0.25 standard deviations (about 25 points) for PISA mathematics when the other variables in the model were held constant. Two SES variables, *home possessions* and *parental education,* showed about the same effect size, with standardized coefficients of 0.11 ($p < .001$). Thus, there was 0.11 standard deviations increase (about 11 points) for PISA 2012 mathematics per one unit of standard deviation increase in students' *home possessions* or *parental education* when the other variables were held constant. *Grade repetition* (−0.19, $p < .001$) and *mathematics anxiety* (−0.17, $p < .001$) were negative predictors of mathematics, as expected. Thus, it can be seen that three non-cognitive variables, *mathematics self-efficacy*, *grade repetition* and *mathematics anxiety,* showed stronger effect sizes as individual-level predictors of mathematics than the two SES-related variables, *home possessions* and *parental education*. The fixed effect of *mathematics self-concept* (0.05, $p < .001$) was much smaller than, and perhaps in the presence of, other self-belief constructs (*mathematics self-efficacy* and *mathematics anxiety*) in the model.

At Level 2, *home possessions* was first entered to account for the country mean differences in mathematics. The Level 2 variable selection process was guided by the HLM results with the TIMSS data and also based on the between-country level correlations of *home possessions* ($r = 0.521$ in PISA 2012 and $r = 0.763$ in PISA 2003). In the next step, the remaining variables were entered, using backward elimination, as Level 2 predictors of the HLM analysis to explain the country mean differences. The final model, with all seven individual-level and two country-level variables being significant at their respective levels ($p < .001$), was constructed to explain the PISA mathematics at both levels.

Of the two Level 2 variables, *home possessions* was the weaker predictor of mathematics performance (0.28, $p < .001$) than *mathematics self-concept* (−0.49, $p < .001$) in PISA 2012. These fixed effects coefficients indicate that the countries performed about 0.28 standard deviations (about 28 points) higher or about 0.49 standard deviations (about 49 points) lower in mathematics per one standard deviation unit increase in *home possessions* or *mathematics self-concept*, respectively, when the other predictors in the model were held constant. The finding of opposite signs in the fixed effect coefficients of the Level 2 predictors, i.e., *home possessions* (i.e., positive) and *mathematics self-concept* (i.e., negative), is consistent with those obtained in the predictability of the Level 2 predictors, *home possessions* (i.e., positive) and *confidence in mathematics* (i.e., negative), in the TIMSS 2011, 2007, and 2003 data.

3.7.4. Evidence from PISA 2003

The steps taken with the PISA 2012 data were also applied to the PISA 2003 data for HLM model construction. As can be seen in Table 5, the results of both years are similar to each other. While the variable selection processes for the HLM analysis were carried out separately for the PISA 2012 and 2003 data, the same sets of variables were identified as the best predictors at the individual and country levels, and even the fixed effect sizes of the same variables were very close to each other in the 2012 and 2003 datasets. For example, the fixed effect estimate of *mathematics self-efficacy* at Level 1 was the same value of 0.25 ($p < .001$) in 2012 and 2003. The fixed effects of *home possessions* were also close at 0.11 ($p < .001$) in 2012 and 0.13 ($p < .001$) in 2003. The effect sizes of *grade repeat* (in 2012) and *expected educational level of student* in 2003 are similar in strength, but in different directions, as expected (i.e., −0.19, $p < .001$; 0.22, $p < .001$, respectively). At Level 2, *home possessions* and *mathematics self-concept* turned out to be the best predictors of country-level differences in both PISA 2003 and 2012 (as well as in TIMSS 2011, 2007, and 2003). The effect size of *home possessions* (0.57, $p < .001$) at Level 2 was stronger than *mathematics self-concept* (−0.34, $p < .001$) in PISA 2003, but *mathematics self-concept* (−0.49, $p < .001$) was stronger than *home possessions* (0.28, $p < .001$) in PISA 2012.

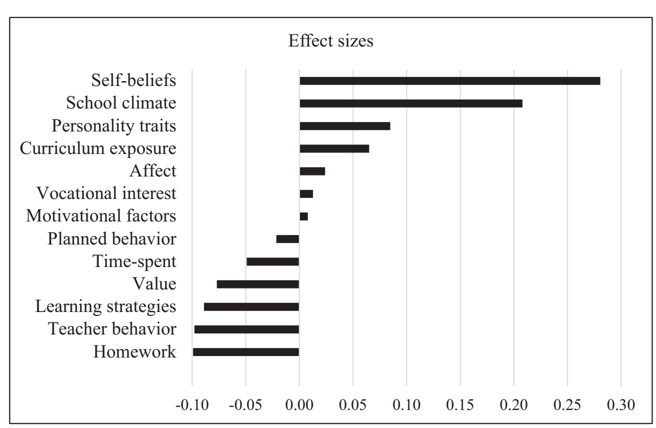

Fig. 1. Effect sizes of TIMSS and PISA non-cognitive constructs classified into research domains.
Note. Effect sizes were calculated as the averages of the TIMSS 2011 and PISA 2012 pan-cultural correlations with students' mathematics scores. The correlations of the previous years were used when the data were not available for particular constructs in the 2011 and 2012 datasets. The constructs included in each domain are listed in Table 1.

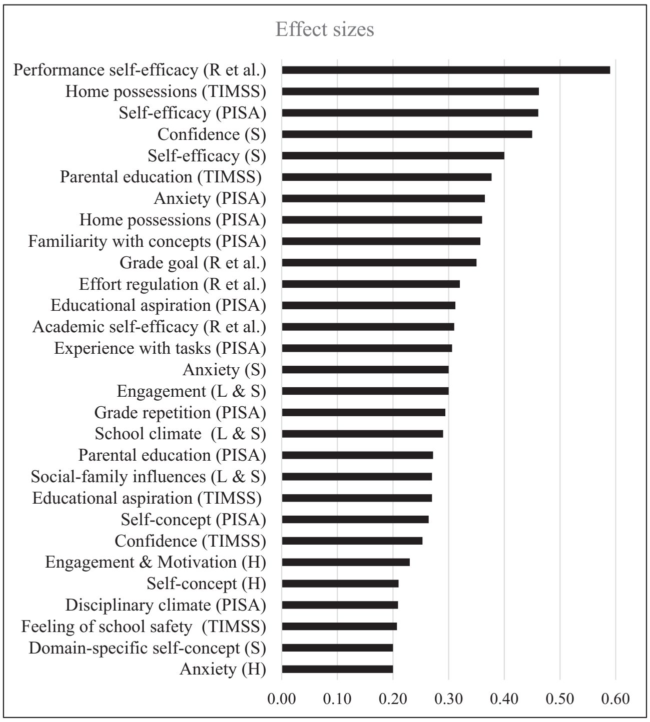

Fig. 2. Effect sizes of non-cognitive and socio-economic status constructs from TIMSS and PISA that showed the absolute values of the pan-cultural correlations > 0.20, and the other major constructs reviewed.
Note. (H) refers to Hattie (2009), (L & S) refers to Lee and Shute (2010), (R et al.) refers to Richardson et al. (2012), and (S) refers to Stankov (2013).

4. Discussion

The purpose of the present study was to identify non-cognitive constructs that best predict students' mathematics achievement based on large-scale international data. The use of international data sources allowed for the identification of non-cognitive variables of 'broad and universal' relevance to student achievement at individual and country levels. We found close similarity in the main findings *between* the TIMSS and PISA surveys as well as *within* the TIMSS and PISA across the years. The convergence of the main findings between the TIMSS and PISA can be taken as quite robust and compelling, given that the two assessments are conducted separately by different organisations, each assessment contains a large number of constructs that do not entirely overlap, student participants were newly sampled for each assessment year, the year gaps between the administrations within the same assessment were 8 and 9 years in TIMSS and PISA, respectively, and the participant countries were not exactly the same between the two assessments and within the same assessment over the years (e.g., 41 countries in PISA 2003 and 64 countries in PISA 2012).

4.1. Synthesis of the findings of the present study at the domain-level

To present a 'big-picture view', we further synthesized the results of our study (Fig. 1) and combined them with the major findings of the previous review papers mentioned in the introduction (Fig. 2). The effect sizes in Fig. 1 were the averages of the pan-cultural correlations of the TIMSS and PISA data (in Tables 2 and 3) across the non-cognitive constructs that were grouped into the same research domains (as presented in Table 1). It comes as no surprise, given the construct-level findings reported in the previous sections, that the effect size of *self-beliefs*, which combined the effect sizes of confidence (TIMSS), self-efficacy (PISA), self-concept (PISA), anxiety (PISA), educational aspiration (TIMSS), and expected educational level of student (PISA), outperformed all the other research domains. This is the only domain that demonstrated a moderately strong effect size (approaching $r = 0.30$). The importance of self-beliefs has been highly advocated in the context of the *social-cognitive theory* by Bandura (1997) as well as by the predictability gradient hypothesis of Stankov (2013) (also see Lee & Stankov, 2013; Pipere & Mieriņa, 2017). The next strongest effect size, just exceeding $r = 0.20$, was found in the domain of *school climate*, which combined the effect sizes from two scales of feeling of school safety (TIMSS) and disciplinary climate (PISA). Learning environment factors, such as school climate, have been emphasized in Lee and Shute (2010) and Wentzel and Watkins (2011). As can be seen in Fig. 1, the remaining domains showed very small, near zero, or negative signs of correlations.

We have also synthesized the effect sizes from the present study based on the absolute values of the pan-cultural correlations in the most recent assessment years (as shown in Tables 2 and 3) and the conclusions drawn about the major non-cognitive constructs in the four review papers—Hattie (2009), Lee and Shute (2010), Richardson et al. (2012) and Stankov (2013). Only the constructs with effect sizes > 0.20 are listed in Fig. 2. Many similarities are noted in the results across the four review papers and the present study utilizing the TIMSS and PISA data. Stankov's (2013) conclusions on the effect sizes of self-efficacy ($r = .40$s) and anxiety ($r = .30$s) are well within the range of the effect sizes reported in PISA's self-efficacy ($r = 0.46$) and anxiety ($r = 0.37$). The TIMSS' confidence ($r = 0.25$) and PISA's self-concept ($r = 0.26$) scales, which adopted a similar set of items, produced very close effect sizes, which were also in similar ranges to the effect sizes of self-concept ($r = .20$s) reported in Hattie (2009) and Stankov (2013). Two school climate scales from TIMSS ($r = 0.207$) and PISA ($r = 0.209$) produced almost identical results. In addition, the effect size of the PISA's educational aspiration ($r = 0.31$) was close to that of the TIMSS' educational aspiration ($r = 0.27$).

Out of 29 constructs listed in Fig. 2, nearly half belong to self-beliefs constructs (e.g., self-efficacy, self-concept, aspiration). Eight of them relate to home SES and school environment factors (e.g., school climate, home possessions, and parental education). The rest can be seen as motivational aspects of learning such as effort regulation and engagement, which were not supported by the findings of the current study.

4.2. Self-beliefs are the strongest predictors of achievement in TIMSS and PISA

In the TIMSS surveys, *confidence in mathematics* and *student's educational aspirations* were the strongest predictors at the individual level while *confidence in mathematics* was also a significant predictor at the country level. In the PISA surveys, *mathematics self-efficacy* was the strongest predictor of achievement in both PISA 2012 and 2003, followed by *mathematics anxiety*, *student's own expected educational level* (only in PISA 2003) and *grade repetition* (only in PISA 2012), at the student individual level. Furthermore, *mathematics self-concept* was one of the two significant predictors of mathematics achievement at the country-level in the PISA data. Overall, *confidence* (in TIMSS) and *self-efficacy* (in PISA) were the strongest predictors of student achievement.

A common and perhaps critical feature of the self-beliefs constructs might be that they seem to tap into projective judgements about self and what an individual can or cannot accomplish in future tasks/events. For educational aspiration, the participants projected themselves as being able to *complete or not complete* a particular level of schooling. For confidence, they were asked to evaluate whether they will "*do well* in mathematics" or if their "teacher thinks that s/he *can do well* in mathematics with difficult materials". Although projective judgements would be influenced by past or current achievement (see Bandura, 1997), it can be inferred that students' *future* projection about their capabilities, rather than the perceptions about their past or current performance, would be strongly associated with their *current* achievement! It is also possible that projective judgements represent the core and common ingredient across different self-beliefs constructs (Lee & Stankov, 2013), symbolizing students' overall tendencies and disposition that emerge through a host of other self-directional behaviors and thoughts such as self-determination, persistence, and willingness to invest their time and effort.

4.3. Non-cognitive versus SES variables

As Fig. 2 shows, the effect sizes of SES variables are not necessarily stronger than those of some of the self-beliefs constructs. In fact, various types of self-efficacy (performance self-efficacy in Richardson et al., 2012; mathematics self-efficacy in PISA; general self-efficacy in Stankov, 2013; academic self-efficacy in Richardson et al., 2012) surpassed the effect size of the parental education variable in the PISA data ($r = 0.27$). Furthermore, our HLM analysis shows stronger effect sizes for *confidence* (TIMSS) and *self-efficacy* (PISA) and *student's educational aspirations* (in TIMSS and PISA) than for the SES measures (*home possessions* and *parental education*) on student achievement at the individual level (see Tables 4 and 5). This was one of the surprising findings of the present study, and therefore provides an empirical basis for future debates over the influence of SES versus non-cognitive constructs on student achievement.

4.4. Self-belief measures beyond the TIMSS and PISA

Given the superior predictability of the self-belief constructs demonstrated in this study, it may be profitable to consider them and other self-related constructs in more detail. Hansford and Hattie (1982) listed the following self-constructs as those that have been extensively studied in educational psychology: *self-concept, self-concept of ability, self-esteem, self-acceptance, self-perception, ideal-self, self-assurance, self-sentiment, self-attitude, self-confidence, self-regard, self-actualisation, identity development* and *self-expectation*. Many more self- or self-related constructs have been employed in recent empirical studies, including *global self-esteem* (Marsh, 1986), *academic self-concept* (Guay, Marsh, & Boivin, 2003), *self-efficacy* (Bandura, 1997), *self-regulation* (Zimmerman, Bandura, & Martinez-Pons, 1992), *self-reliance* (Winne, 2004), *self-discipline* (Duckworth & Seligman, 2005), *academic self-handicapping* (Urdan, 2004), *mental toughness* (see McGeown, St Clair-Thompson, & Clough, 2016), *grit* (Duckworth, Peterson, Matthews, & Kelly, 2007), *academic resilience* (Martin & Marsh, 2009), *locus of control* (Rotter, 1966; Stipek & Weisz, 1981), *cognitive-item performance based confidence* (Stankov & Lee, 2008, 2015) and *attribution of failure* (see Schunk, 1983). Further empirical investigation is required to determine the extent of conceptual overlap across different self-constructs and TIMSS and the PISA's self-belief constructs.

4.5. Too Many non-cognitive constructs do not correlate with achievement in mathematics

Our analysis of 65 non-cognitive variables across the TIMSS and PISA indicates that only three out of 22 non-cognitive measures in the TIMSS and only eight out of 43 non-cognitive measures in the PISA have moderately strong, direct associations with student achievement in mathematics (see Tables 2 and 3). Many non-cognitive variables examined in this study showed low correlations with mathematics achievement. Among the TIMSS variables, these include measures of *positive affect towards mathematics, valuing mathematics, extracurricular time spent in mathematic-type activities*, and *engagement with mathematics lessons*. In addition, broader measures such as *attitude towards school* and *parental involvement* have essentially zero correlation with mathematics achievement. The PISA list of non-cognitive variables that do not correlate with achievement includes similar measures to those in the TIMSS list, but several additional variables are also included in PISA, such as *instrumental and (intrinsic) interest* and *motivations*, different types of *teacher support*, *learning strategies*, *school learning time*, *time spent on mathematics-related activities* including *homework*, and *co-operative and competitive learning style* (see Tables 2 and 3). Other non-cognitive variables traditionally studied in personality research, such as *openness to problem solving* (representing openness to experience, see Ackerman, Chamoro-Pemuzic, & Furnham, 2010) and *work ethic* (representing conscientiousness, see Poropat, 2009), also failed to demonstrate moderately strong correlations with student achievement in the PISA 2012 data. Given that the TIMSS and PISA non-cognitive measures were chosen as the “most” education-relevant constructs, it is surprising that many of them showed near or essentially zero correlations with student achievement.

4.6. Implications for policy debate

The measures that demonstrated at least moderately strong predictive validity for achievement at both individual and country levels (i.e., *confidence* in the TIMSS and *self-concept* in the PISA, and *home possessions* in both TIMSS and PISA) may be particularly important because the findings may be linked to various conditions at the individual as well as system/country-level. Traditionally, student SES background has been viewed as “fixed” and exerting indirect influences on student achievement. Thus, interventions based on SES are largely considered to be outside of the paradigms of school-based programs. However, it is not entirely irrational to promote SES-related variables in the school system, for instance, by advocating the importance of education of parents themselves in addition to their children, or by implementing the systematic evaluation of the availability of adequate or minimum-level education-related resources at home (e.g., books and space to study). As our within-country and between-country analyses suggest, such provision would be more important for less-developed countries. School-based initiatives may also be designed to provide appropriate monitoring and assessment of student self-belief in their academic work. Intervention programs targeting a good calibration of self-beliefs and reduction in test anxiety might be particularly useful in countries having lower academic achievement scores (see Lee, 2009). Supporting policy measures to help students to become less anxious and appropriately confident in school work could be planned at the system level (OECD, 2015, p. 2).

4.7. Limitations of the present study

The present study has a few obvious limitations. First, the findings reported herein are based on correlations and no methodology was used to examine causal effects. It is possible that the effect sizes reported in this paper are the outcomes of bi-directional cyclic relationships between the constructs. High predictability of self-beliefs in mathematics can also mean that students who are better in mathematics simply know that they are better in mathematics, which also signals a link to metacognition (i.e., knowing what they know). Second, the questionnaire data were based on self-report. While we acknowledge that some discrepancy may exist between student reporting and actual behaviors in learning habits and practices (e.g., Pape, Bell, & Yetkin, 2003), no validation methods were adopted to confirm the “truthfulness” of students' self-reports. Third, we did not venture out to explore the reasons behind the many unexpected, negative results of this study. In fact, too many non-cognitive constructs failed to show even a medium effect size. While the investigation (or even speculation) of the negative findings is beyond the scope of the present study, its findings may contribute to decision-making about non-cognitive constructs that are likely to produce a positive relationship to student learning.

References

Ackerman, P. L., Chamoro-Pemuzic, T., & Furnham, A. (2010). Trait complexes and academic achievement: Old and new ways of examining personality in educational contexts. *British Journal of Educational Psychology, 81*, 27–40. http://dx.doi.org/10.

1348/000709910X522564.
Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology, 22*, 453–474.
Alexander, P. A. (1992). Domain knowledge: Evolving issues and emerging concerns. *Educational Psychologist, 27*, 33–51. http://dx.doi.org/10.1207/s15326985ep2701_4.
Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject matter knowledge affects recall and interest. *American Educational Research Journal, 31*, 313–337. http://dx.doi.org/10.3102/00028312031002313.
Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.
Bempechat, J. (2004). The motivational benefits of homework: A social cognitive perspective. *Theory Into Practice, 43*, 189–196. http://dx.doi.org/10.1207/s15430421tip4303_4.
Bowles, S., & Gintis, H. (1976). *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life.* New York: Basic Books.
Brown, S. D., & Lent, R. W. (2015). Social cognitive career theory: A theory of self (efficacy) in context. In F. Guay, D. McInerney, R. Craven, & H. Marsh (Eds.). *Self-concept, motivation and identity: Underpinning success with research and practice* (pp. 173–200). Charlotte, NC: Information Age Publishing.
Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*, 778–820. http://dx.doi.org/10.3368/jhr.XLI.4.778.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research, 76*, 1–62. http://dx.doi.org/10.3102/00346543076001001.
Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica, 78*(3), 883–931.
Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939–944. http://dx.doi.org/10.1111/j.1467-9280.2005.01641.x.
Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*, 237–251. http://dx.doi.org/10.3102/0013189X15584327.
Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087–1101. http://dx.doi.org/10.1037/0022-3514.92.6.1087.
Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 3*, 215–225. http://dx.doi.org/10.1177/0146167295213003.
Elliott, M. A., Armitage, C. J., & Baughan, C. J. (2003). Drivers' compliance with speed limits: An application of the theory of planned behavior. *Journal of Applied Psychology, 88*, 964–972. http://dx.doi.org/10.1037/0021-9010.88.5.964.
Eshel, V., & Kohavi, R. (2003). Perceived classroom control, self-regulated learning strategies, and academic achievement. *Educational Psychology, 23*, 249–260. http://dx.doi.org/10.1080/0144341032000060093.
Farkas, G. (2003). Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual Review of Sociology, 29*, 541–562. http://dx.doi.org/10.1146/annurev.soc.29.010202.100023.
Foy, P., Brossman, B., & Galia, J. (2012). *Scaling the TIMSS and PIRLS 2011 achievement data.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109. http://dx.doi.org/10.3102/00346543074001059.
Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*, 124–136. http://dx.doi.org/10.1037/0022-0663.95.1.124.
Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and academic performance measures. *Review of Educational Research, 52*, 123–142. http://dx.doi.org/10.3102/00346543052001123.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses related to achievement.* London: Routledge, Taylor and Francis Group.
Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99–136. http://dx.doi.org/10.3102/00346543066002099.
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482. http://dx.doi.org/10.1086/504455.
Holden, G. W., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: A meta-analysis. *Psychological Reports, 66*, 1044–1046. http://dx.doi.org/10.2466/pr0.1990.66.3.1044.
Hoy, W. K., & Hannum, J. W. (1997). Middle school climate: An empirical assessment of organizational health and student achievement. *Educational Administration Quarterly, 33*, 290–311. http://dx.doi.org/10.1177/0013161X97033003003.
Hunley, S. A., Evans, J. H., Delgado-Hachey, M., Krise, J., Rich, T., & Schell, C. (2005). Adolescent computer use and academic achievement. *Adolescence, 40*, 307–318.
Joncas, M., & Foy, P. (2012). *Sample design in TIMSS and PIRLS.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Jussim, L., Robustelli, S., & Cain, T. (2009). Teacher expectancies and self-fulfilling prophecies. In K. R. Wentzel, & A. Wigfield (Eds.). *Handbook of motivation at school* (pp. 349–379). New York, NY: Routledge/Taylor and Francis Group.
Kleiber, D. A., Larson, R., & Csikszentmihalyi, M. (1986). The experience of leisure in adolescence. *Journal of Leisure Research, 18*, 169–176.
Kyllonen, P. C., & Bertling, J. P. (2013). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.). *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 277–285). London, England: Chapman & Hall.
Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*, 355–365. http://dx.doi.org/10.1016/j.lindif.2008.10.009.
Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in K–12 academic performance: An integrative perspective on student learning. *Educational Psychologist, 45*, 185–202. http://dx.doi.org/10.1080/00461520.2010.493471.
Lee, J., & Stankov, L. (2013). Higher-order structure of motivation, self-beliefs, learning strategies, and attitudes toward school and its prediction of PISA 2003 mathematics scores. *Learning and Individual Differences, 26*, 119–130. http://dx.doi.org/10.1016/j.lindif.2013.05.004.
Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*, 79–122. http://dx.doi.org/10.1006/jvbe.1994.1027.
Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review, 18*, 307–314. http://dx.doi.org/10.1007/s10648-006-9028-x.
Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology, 51*, 1224–1236. http://dx.doi.org/10.1037/0022-3514.51.6.1224.
Martin, A. J., & Marsh, H. W. (2009). Academic resilience and academic buoyancy: Multidimensional and hierarchical conceptual framing of causes, correlates and cognate constructs. *Oxford Review of Education, 35*, 353–370. http://dx.doi.org/10.1080/03054980902934639.
Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (Eds.). (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: Boston College.
Martin, M. O., Mullis, I. V. S., Foy, P., & Arora, A. (2012). *Creating and interpreting the TIMSS and PIRLS 2011 Context Questionnaire Scales.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
McGeown, S. P., St Clair-Thompson, H., & Clough, P. (2016). The study of non-cognitive attributes in education: Proposing the mental toughness framework. *Educational Review, 68*, 96–113. http://dx.doi.org/10.1080/00131911.2015.1008408.
Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Educational Psychology, 71*, 281–292. http://dx.doi.org/10.1037/0022-0663.71.3.281.
Messick, S. (1996). Bridging cognition and personality in education: The role of style in performance and development. *European Journal of Personality, 10*, 353–376. http://dx.doi.org/10.1002/j.2333-8504.1996.tb01700.x.
Mickelson, R. A. (1990). The attitude-achievement paradox among black adolescents. *Sociology of Education, 63*, 44–61. http://dx.doi.org/10.2307/2112896.
Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counselling Psychology, 38*, 30–38. http://dx.doi.org/10.1037/0022-0167.38.1.30.
OECD (2004). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003.* OECD Publishing.
OECD (2005). *PISA 2003 technical report.* OECD Publishing.
OECD (2011). *PISA in focus: How do some students overcome their socio-economic background?* OECD Publishing.
OECD (2012). *PISA 2009 technical report.* PISA, OECD Publishinghttp://dx.doi.org/10.1787/9789264167872-en.
OECD (2013). *PISA 2012 results: Ready to learn: Students' engagement, drive and self-beliefs (volume III).* PISA: OECD Publishinghttp://dx.doi.org/10.1787/9789264201170-en.
OECD (2014). *PISA 2012 technical report.* OECD Publishing.
OECD (2015). *PISA in focus: How confident are students in their ability to solve mathematics problems?* OECD Publishing.
Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report.* Chestnut Hill, MA: Boston College.
Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology, 78*, 569–582. http://dx.doi.org/10.1037/0021-9010.78.4.569.
Pape, S. J., Bell, C. V., & Yetkin, I. E. (2003). Developing mathematical thinking and self-regulated learning: A teaching experiment in a seventh-grade mathematics classroom. *Educational Studies in Mathematics, 53*, 179–202.
Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99*, 83–98. http://dx.doi.org/10.1037/0022-0663.99.1.83.
Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40. http://dx.doi.org/10.1037/0022-0663.82.1.33.
Pipere, A., & Mieriņa, I. (2017). Exploring non-cognitive predictors of mathematics achievement among 9th grade students. *Learning and Individual Differences, 59*, 65–77. http://dx.doi.org/10.1016/j.lindif.2017.09.005.
Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322–338. http://dx.doi.org/10.1037/a0014996.
Raudenbush, S. W., Bryk, A. S., Cheong, A. S., Fai, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modelling.* Lincolnwood, IL: Scientific Software International.
Reynolds, A. J., & Walberg, H. J. (1992). A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education, 23*, 306–328. http://dx.doi.org/10.2307/749308.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. http://dx.doi.org/10.1037/a0026838.
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.
Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28.
Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*, 848–856. http://dx.doi.org/10.1037/0022-0663.75.6.848.
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417–453. http://dx.doi.org/10.3102/00346543075003417.
Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modelling* (2nd ed.). Thousand Oaks, CA: Sage.
Stankov, L. (2011). Individual, country and societal cluster differences on measures of personality, attitudes, values and norms. *Learning and Individual Differences, 21*, 55–66. http://dx.doi.org/10.1016/j.lindif.2010.09.002.
Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences, 55*, 727–732. http://dx.doi.org/10.1016/j.paid.2013.07.006.
Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*, 961–976. http://dx.doi.org/10.1037/a0012546.
Stankov, L., & Lee, J. (2014). Quest for the best noncognitive predictor of academic achievement. *Educational Psychology, 34*, 1–8. http://dx.doi.org/10.1080/01443410.2013.858908.
Stankov, L., & Lee, J. (2015). Confidence and self-efficacy: Are they different? In F. Guay, D. McInerney, R. Craven, & H. Marsh (Eds.). *Self-concept, motivation and identity: Underpinning success with research and practice* (pp. 225–248). Charlotte, NC: Information Age Publishing.
Stankov, L., & Lee, J. (2016). Nastiness, morality and religiosity in 33 nations. *Personality and Individual Differences, 99*, 56–66. http://dx.doi.org/10.1016/j.paid.2016.04.069.
Stankov, L., Lee, J., & von Davier, M. (2017). A note on construct validity of the anchoring method in PISA 2012. *Journal of Psychoeducational Assessment*. http://dx.doi.org/10.1177/0734282917702270.
Stipek, D. J., & Weisz, J. R. (1981). Perceived personal control and academic achievement. *Review of Educational Research, 51*, 101–137. http://dx.doi.org/10.3102/00346543051001101.
Turner, R., & Adams, R. J. (2007). The programme for international student assessment: An overview. *Journal of Applied Measurement, 8*, 237–248.
Urdan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology, 96*, 251–264. http://dx.doi.org/10.1037/0022-0663.96.2.251.
Van de Gaer, E., Grisay, A., Schulz, 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, 43*, 1205–1228. http://dx.doi.org/10.1177/0022022111428083.
Voelkl, K. E. (2012). School identification. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.). *Handbook of research on student engagement* (pp. 193–218). New York, NY: Springer.
Ware, H., & Kitsantas, A. (2007). Teacher and collective efficacy beliefs as predictors of professional commitment. *Journal of Educational Research, 100*, 303–310. http://dx.doi.org/10.3200/JOER.100.5.303-310.
Wentzel, K. R. (2009). Students' relationships with teachers as motivational contexts. In K. R. Wentzel, & A. Wigfield (Eds.). *Handbook of motivation at school* (pp. 301–322). New York, NY: Routledge/Taylor and Francis Group.
Wentzel, K. R. (2015). Competence within context: Implications for the development of positive student identities and motivation at school. In F. Guay, D. McInerney, R. Craven, & H. Marsh (Eds.). *Self-concept, motivation and identity: Underpinning success with research and practice* (pp. 299–336). Charlotte, NC: Information Age Publishing.
Wentzel, K. R., & Watkins, D. E. (2011). Peer relationships and learning: Implications for instruction. In R. Mayer, & P. Alexander (Eds.). *Handbook of research on learning and instruction* (pp. 322–343). New York, NY: Routledge.
Wentzel, K. R., Russell, S., & Baker, S. A. (2016). Emotional support and expectations from parents, teachers, and peers predict adolescent competence at school. *Journal of Educational Psychology, 108*, 242–255. http://dx.doi.org/10.1037/edu0000049.
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68–81. http://dx.doi.org/10.1006/ceps.1999.1015.
Wigfield, A., Cambria, J., & Eccles, J. S. (2012). Motivation in education. In R. M. Ryan (Ed.). *The Oxford handbook of motivation* (pp. 463–478). New York, NY: Oxford University Press.
Winne, P. H. (2004). Putting volition to work in education. *Teachers College Record, 106*, 1879–1887.
Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 21*, 3–17. http://dx.doi.org/10.1207/s15326985ep2501_2.
Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*, 663–676.