Journal of Educational Psychology
2017, Vol. 109, No. 3, 439–449

0022-0663/17/$12.00 http://dx.doi.org/10.1037/edu0000140

Academic Competencies: Their Interrelatedness and Gender Differences at Their High End

Sebastian Bergold, Heike Wendt, Daniel Kasper, and Ricarda Steinmayr
Technical University Dortmund

The present study investigated (a) how a latent profile analysis based on representative data of N = 74,868 4th graders from 17 European countries would cluster the students on the basis of their reading, mathematics, and science achievement test scores; (b) whether there would be gender differences at various competency levels, especially among the top performers; (c) and whether societal gender equity might account for possible cross-national variation in the gender ratios among the top performers. The latent profile analysis revealed an international model with 7 profiles. Across these profiles, the test scores of all achievement domains progressively and consistently increased. Thus, consistent with our expectations, (a) the profiles differed only in their individuals' overall performance level across all academic competencies and not in their individuals' performance profile shape. From the national samples, the vast majority of the students could be reliably assigned to 1 of the profiles of the international model. Inspection of the gender ratios revealed (b) that boys were overrepresented at both ends of the competency spectrum. However, there was (c) some cross-national variation in the gender ratios among the top performers, which could be partly explained by women's access to education and labor market participation. The interrelatedness of academic competencies and its practical implications, the role of gender equity as a possible cause of gender differences among the top performers, and directions for future research are discussed.

Keywords: academic achievement, gender differences, TIMSS and PIRLS, gender equity, latent profile analysis

Both in educational research and practice and by students themselves, the belief prevails that most students exhibit considerable strengths in some domains (e.g., reading) and at the same time considerable weaknesses in other domains (e.g., mathematics; Marsh & Hau, 2004; Wang, Eccles, & Kenny, 2013). In stark contrast to this belief stand empirical findings documenting that competencies in different domains such as reading, mathematics, and science are highly intercorrelated (e.g., Reilly, 2012; Rindermann, 2007). These findings imply that higher competencies in one domain are likely to be accompanied by higher competencies in the other domains. Studies that investigate student profiles across different academic competency domains are lacking, even though studies using an intraindividual approach provide a more comprehensive picture of students' overall academic competency than studies only focusing on one or two competencies (e.g., Brunner et al., 2013). Due to the high intercorrelations among different competencies, we hypothesized that students' competency profiles would only differ in the absolute values across competency domains but not in the shape of different competencies relative to each other when investigating reading, mathematics, and science competencies. Academic competencies in different domains are partly shaped by the numerous determinants located at the country, the school, the classroom, and the student level (e.g., Byrnes & Miller, 2007). As these processes seem to work equally in comparable schooling systems, we expected that the students' profile patterns could be replicated between countries with comparable schooling systems.

Moreover, we examined gender differences in the competency profiles. We were especially interested in the profile representing those students having the highest competencies across all domains. Previous research has found that boys were overrepresented in the upper tail of the mathematics and science distribution but girls in reading competencies (e.g., Hedges & Nowell, 1995; Nowell & Hedges, 1998). Hitherto, the question has been unanswered whether more boys or girls are present in the right tail of the ability distribution when considering all three competencies simultaneously. Answering this question is important for the ongoing debate on gender differences in academic competencies and on women's underrepresentation in scientific careers (Ceci, Williams, & Barnett, 2009; Hyde, 2014). Although research suggests that gender differences in payment and promotion opportunities as well as women's interests, career preferences, and variety of choice options considerably contribute to women's underrepresentation in science (Ceci et al., 2009; Ferriman, Lubinski, & Benbow, 2009; Hunt, 2016; Wang et al., 2013), high competencies in academic domains such as reading, mathematics, and science (and spatial ability; e.g., Wai, Lubinski, & Benbow, 2009) are a further im-

This article was published Online First July 18, 2016.

Sebastian Bergold, Department of Psychology, Technical University Dortmund; Heike Wendt and Daniel Kasper, Institute for School Development Research, Technical University Dortmund; Ricarda Steinmayr, Department of Psychology, Technical University Dortmund.

Correspondence concerning this article should be addressed to Sebastian Bergold, Department of Psychology, Technical University Dortmund, Emil-Figge-Straße 50, D-44227 Dortmund, Germany. E-mail: sebastian.bergold@tu-dortmund.de

portant prerequisite for scientific careers (Ceci et al., 2009). Furthermore, it is important to investigate whether the gender differences in the highest competency profile are related to any cultural specifics of the countries we considered. Doing so serves to shed further light on the reasons for why males or females are overrepresented at the highest end of mathematics, reading, and science performance.

In the present study, we used representative data of elementary school children from the Trends in International Mathematics and Science Study (TIMSS) 2011 and the Progress in International Reading Literacy Study (PIRLS) 2011 from 17 European countries that were members of the European Union in 2011. Both studies were run simultaneously and the same children were investigated with the competency tests from both TIMSS and PIRLS. Applying a person-centered approach, we investigated (a) whether a latent profile analysis (LPA) on the basis of mathematics, science, and reading achievement test scores would reveal profiles differing in their individuals' overall performance level rather than in their individuals' performance profile shape; (b) whether there are gender differences in any of the profiles and especially in the profile representing the highest performance level; and (c) whether we would find cross-national variability of the gender ratios in the highest competency profile and whether this variability might be explained by societal gender equity.

Toward a Comprehensive Perspective on Academic Competencies

Different studies find academic achievement tests to be highly intercorrelated. For example, the reading, mathematics, and science scores from PISA were found to intercorrelate from $r = .75$ to $r = .81$ (Reilly, 2012) or even from $r = .95$ to $r = .99$ (Rindermann, 2007). In TIMSS and PIRLS 2011, the correlations were lower but still substantial, ranging from $r = .54$ to $r = .74$ (Bos et al., 2012). Students doing better than other students in, for example, reading do on average also better in mathematics and science. This is because mathematics and science tasks might to a certain degree require reading skills or because reading, mathematics, and science might all be influenced by superordinate factors.

Indeed, children's academic competencies are the results of a myriad of determinants located at the country, the school, the classroom, and the student level (see, e.g., Byrnes & Miller, 2007). A lot of these shape interindividual differences in students' academic achievement in a rather uniform way, that is, in about the same manner across different domains. This might explain why countries achieving certain levels in one domain tend to achieve similar levels also in other domains (e.g., Martin, Mullis, Foy, & Stanco, 2012; Mullis, Martin, Foy, & Arora, 2012; Mullis, Martin, Foy, & Drucker, 2012; Organization for Economic Cooperation and Development [OECD], 2014). Even though these country-level differences are not yet well understood, some explanations are discussed that refer to factors influencing the entire educational system, for example, differences in the states' investments in education (e.g., OECD, 2014). Furthermore, variables located at the school and at the classroom level (see Chung, 2015; Kellaghan, 2015) or—highly correlated with indices of school quality—students' family background variables, such as the socioeconomic status (SES), are correlated with performance in different domains (see Kurtz-Costes, 2015; OECD, 2014; Sirin, 2005). The same is true for student-level determinants. At the student level, the most important single determinant of academic achievement is cognitive ability, which is highly correlated with competencies in all domains (e.g., Deary, Strand, Smith, & Fernandes, 2007; Jensen, 1998). Cognitive ability is not only associated with school performance but also with variables influencing school success at different levels such as family background or school environment (see, e.g., Chung, 2015). In addition, motivational variables and school-relevant personality variables such as conscientiousness, openness for experience, or test anxiety are also important predictors of academic achievement across domains and are correlated with other determinants on the same and other levels (e.g., Hembree, 1988; Steinmayr, Dinger, & Spinath, 2010, 2012; Steinmayr & Spinath, 2009).

Summing up, there are several, partly interrelated, variables having a general influence on academic competencies across domains that might explain the high correlations between competencies in different domains. As they are found across countries with comparable schooling systems, we assume that they impact on students' competencies all in the same way. More precisely, when investigating a large sample comprising students from a wide range of different settings—homes, classes, schools, and countries—one would expect that students doing well in one domain should also tend to do well in other domains. Students should have rather balanced ability profiles instead of unbalanced profile shapes, and this should hold across the countries.

For the investigation of ability profiles, a person-centered approach should be the method of choice. However, to our knowledge, so far only one study has used such an approach. Wang et al. (2013) conducted a LPA with SAT reading and mathematics scores from 1,490 12th graders. They identified five profiles: high math/high verbal scores ($n = 298$), high math/moderate verbal scores ($n = 373$), moderate math/moderate verbal scores ($n = 402$), low math/high verbal scores ($n = 298$), and low math/low verbal scores ($n = 119$). At first glance, these results seem partly to contradict the expectation of balanced ability profiles. However, Wang et al.'s (2013) sample was preselected for ability level, in that it was composed of "intellectually able, college-bound U.S. students" (p. 771). In samples of preselected and more abled individuals, the influence of important cross-domain determinants such as family background or school quality might vanish. Furthermore, because of the preselection, Wang et al.'s (2013) terms *high*, *moderate*, and *low scores* have to be seen in relation to the ability group they investigated. For example, a value of 655 points (on the SAT subtest scale from 200 to 800 points with an average of 500 points) was termed *moderate* (p. 772). However, when compared with scores that a population-representative sample would obtain, 655 points would reflect an above-average result, just like the scores termed *high*. Thus, when bearing the full ability range in mind, the discrepancies between the domains were not as large as the titles of the profiles might have suggested. Moreover, in high ability ranges, narrow abilities often gain in importance relatively to broad abilities (e.g., Reynolds, Keith, & Beretvas, 2010). In addition, the reliability of the measurement often shrinks as the ability level becomes more extreme, especially at the subtest level. This leads to more heterogeneous ability profiles than would be found in a population-representative sample (Rost, 2013). In our study, we relied on representative and large samples from different countries, comprising students from a wide range of homes, classrooms, schools, and countries. Furthermore, we also considered competencies in science in addition to competencies in reading and mathematics.

Gender Differences on Academic Performance Tests

Academic competencies—in particular, the core competencies reading, mathematics, and science—predict a variety of important life outcomes, such as subsequent school grades, participation and academic success in higher education, and success on the labor market (e.g., Kuncel, Hezlett, & Ones, 2001; Richardson, Abraham, & Bond, 2012). Thus, the question whether there are gender differences at different levels of academic competencies is of high practical relevance. This is particularly true for gender differences at the high end of academic competencies because these are likely to contribute to the explanation of why there are far more men than women among the top performers in key societal positions such as in research (Ceci et al., 2009).

A substantial body of research based on large representative samples and a number of meta-analyses have already investigated gender differences on mathematics, science, and reading achievement tests. Some of them have addressed gender differences in mean scores only. However, gender ratios in the tails of the ability distribution can be due to both mean and variance differences between genders (we assume normally distributed scores for each gender; e.g., Feingold, 1992; Hedges & Friedman, 1993). Thus, mean differences, variance differences, and resulting gender ratios in the upper and in the lower tail as well as in the middle range of mathematics, science, and reading ability distributions, respectively, will be reviewed in the following.

Analyses of gender differences in mean mathematics test scores have revealed at most a small effect in favor of boys (e.g., Else-Quest, Hyde, & Linn, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Hedges & Nowell, 1995; Reilly, Neumann, & Andrews, 2015; but see Brunner, Krauss, & Kunter, 2008). Inspection of the variance ratios (i.e., male variance divided by female variance) showed that boys displayed higher variance on math scores than girls (e.g., Lindberg, Hyde, Petersen, & Linn, 2010; Reilly et al., 2015). Variance ratios (VRs) were only small in most cases and typically ranged from 1.05 to 1.20, that is, boys were 5% to 20% more variable than girls. However, greater male variance resulted in a considerable overrepresentation of boys in the upper tail of the mathematics score distribution, all the more if it was combined with a small mean difference in favor of boys. For example, Nowell and Hedges (1998) found that a d of 0.09 and a VR of 1.13 resulted in a gender ratio of 1.62:1 in favor of boys at the top 5%, of 2:1 at the top 3%, and of 2.62:1 at the top 1% of math achievers. Because of boys' slightly greater mean, both in the lower tail and in the middle range of the distribution, there were either no gender differences or girls were slightly overrepresented (Hedges & Nowell, 1995; Machin & Pekkarinen, 2008; Nowell & Hedges, 1998). Similar results were also found for science (e.g., Hedges & Nowell, 1995; Nowell & Hedges, 1998; Reilly et al., 2015).

Whereas studies revealed a substantial overrepresentation of boys among high mathematics and science achievers, they showed the opposite pattern for high reading achievers. For instance, Nowell and Hedges (1998) found girls to outscore boys on mean reading achievement by $d = 0.23$ in 1992. Although boys again tended to have a greater variability than girls (VR = 1.14), the considerable mean difference resulted in a female overrepresentation in the upper tail of the reading ability distribution (see also Machin & Pekkarinen, 2008; Reilly, 2012). Because of boys' greater variability, however, the gender ratio in favor of girls did not increase but decreased with a stricter cut-off criterion (top 5%: 1.33:1; top 3%: 1.27:1; top 1%: 1.11:1). In the lower tail of the distribution, boys were clearly overrepresented (lowest 10%: 1.85; lowest 5%: 1.97), whereas they were slightly underrepresented in the middle range (see also Hedges & Friedman, 1993; Machin & Pekkarinen, 2008).

Thus, gender differences vary depending on the ability level and domains investigated. No study so far has examined gender differences among top achievers considering more than one domain. Doing so is important because all of the academic core competencies are predictors of future success. With the present study, we seek to close this research gap, and investigate whether and, if so, to which extent boys are overrepresented among the highest achieving students when reading competencies (in which girls prevail among the highest achievers) were considered in addition to mathematics and science competencies.

What Causes Gender Differences on Academic Performance Tests?

The reasons for gender differences in academic achievement are not yet well understood. Several explanations have been suggested, which can roughly be subdivided into the biological and the sociocultural account. The biological account encompasses evolutionary, genetic, hormonal, and brain-related explanations. However, evidence for biological theories is mixed and studies with sufficient numbers of highly abled individuals are missing (see, e.g., Ceci et al., 2009; Hyde, 2014, for an overview).

Sociocultural theory posits that gender differences are driven by social influences such as societal gender equity (e.g., equity in labor division, women's access to education; see Ceci et al., 2009; Hyde, 2014). Although the results reviewed above have indicated overall greater means and variability for boys in mathematics and science as well as greater means for girls in reading, studies with large representative international samples have also found considerable variability in gender differences across countries (as well as across ethnicities and cohorts; e.g., Else-Quest et al., 2010; Guiso et al., 2008; Penner, 2008; Reilly, 2012). Consequently, cross-national variability in both mean and variance differences resulted in cross-national variability in gender ratios in the upper tail of the ability distributions, which are especially important with regard to gender differences among the later top performers in society. For example, Guiso et al. (2008) showed that, for every boy at the top 5% mathematics achievers, there were, on average, 0.6 girls. However, this ratio ranged from 0.4 (Korea) to 1.1 (Indonesia; see also Penner, 2008). This variability in findings suggests that the sociocultural context is a crucial factor related to gender differences in achievement test scores. In most cases, gender differences in mathematics and science tended to be smaller, and gender differences in reading tended to be greater, in countries with higher gender equity (Else-Quest et al., 2010; Guiso et al., 2008; Penner, 2008; Reilly, 2012).

However, the studies reported above have examined each academic competency separately. Yet, all three academic competencies are predictors of important life outcomes and are substantially intercorrelated. Therefore, in the evaluation of gender differences, they should be considered simultaneously. We are not aware of any study that pursued this target applying a person-centered approach.

Hypotheses and Research Questions

Evidence from research on academic achievement suggests that different academic competencies are substantially interrelated. Thus, we expected (a) that, in a sample of students from a wide range of settings, we would find student profiles that differ in their overall performance level across all three domains mathematics, science, and reading rather than in the shape of their performance profile. Because the intercorrelations among academic domains are observed internationally (Brunner et al., 2013; Reilly, 2012; Rindermann, 2007); we expected (b) the profile solution to be valid also for the individual countries. We also examined (c) the gender ratios in the different profiles, especially whether boys would be overrepresented among the top performers. To gain insight into the mechanisms underlying the genesis of unequal gender ratios at the high end of academic competencies, we examined (d) whether the gender ratios in the highest profile would vary across the countries. If this would be the case, then we expected (e) that this variation could at least partly be explained by societal gender equity.

Method

Sample

We used representative data from the TIMSS 2011 and the PIRLS 2011. Assessments for TIMSS and PIRLS are usually conducted independently of each other, executed in different cycles and focusing on different academic competencies. In 2011, however, TIMSS and PIRLS coincided with each other for the first time and were thus in some countries conducted mutually, collecting mathematics, science, and reading test data from the same students.

Overall, 32 countries participated in the combined TIMSS 2011 and PIRLS 2011 assessments with nationally representative samples of fourth graders. To foster Europe's economic growth, in 2010 the European Commission launched Europe 2020, an economic program that also entailed the Education and Training 2020 program. Within this program, a common strategic educational framework was set up for all members of the European Union to support them in further developing their educational systems to promote lifelong learning and achieve higher school completion and employment rates as well as greater educational justice (European Commission, 2015). Thus, the member states of the European Union follow the same broad educational framework and have relatively comparable educational systems, at least with regard to elementary school. We aimed to take advantage of this fact because, if cross-national differences were observable, these could not be explained by differences between educational systems. This provides ideal prerequisites to examine the role of societal gender equity in the genesis of gender differences. Therefore, we used the data from the $k = 17$ countries that both took part in the combined TIMSS/PIRLS 2011 assessments and were members of the European Union when data collection took place (i.e., Austria, Czech Republic, Finland, Germany, Hungary, Ireland, Italy, Lithuania, Malta, Northern Ireland, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, and Sweden). The 17 national samples resulted in a total sample of $N = 74{,}868$ fourth-grade students (36,655 girls and 38,213 boys; see Table 1) from 2,704 schools. Due to the sampling procedure implemented in TIMSS and PIRLS, it was ensured that the children in the different countries were all

Table 1
Students' N, Mean Age, and Mean Scores (Standard Errors) in Reading, Mathematics, and Science

Country	*N*			Mean age	Reading		Mathematics		Science	
	Girls	Boys	Total		Girls	Boys	Girls	Boys	Girls	Boys
Austria	2,232	2,355	4,587	10.3	533 (2.2)*	525 (2.7)	504 (2.7)	513 (3.3)*	525 (2.8)	538 (3.6)*
Czech Republic	2,159	2,274	4,433	10.3	549 (2.5)*	542 (2.5)	505 (2.8)	516 (2.7)*	529 (2.9)	544 (2.7)*
Finland	2,223	2,318	4,541	10.8	578 (2.3)*	558 (2.2)	542 (2.5)	549 (2.9)*	570 (2.9)	570 (3.0)
Germany	1,940	1,988	3,928	10.4	545 (2.3)*	537 (2.7)	523 (2.7)	532 (2.6)*	522 (3.0)	534 (3.2)*
Hungary	2,533	2,616	5,149	10.6	547 (3.2)*	532 (3.2)	514 (3.6)	517 (3.9)	532 (4.0)	537 (3.9)
Ireland	2,165	2,218	4,383	10.3	559 (2.9)*	544 (3.0)	526 (3.7)	529 (3.3)	516 (4.0)	516 (4.6)
Italy	2,067	2,058	4,125	9.7	543 (2.4)	540 (2.7)	503 (3.1)	512 (2.9)*	520 (3.2)	528 (3.0)*
Lithuania	2,200	2,384	4,584	10.7	537 (2.4)*	520 (2.4)	533 (2.6)	534 (2.9)	514 (2.4)	515 (3.0)
Malta	1,694	1,798	3,492	9.8	486 (2.4)*	468 (2.0)	492 (1.6)	499 (2.1)*	443 (2.2)	449 (2.8)*
Northern Ireland	1,717	1,752	3,469	10.4	567 (2.5)*	550 (3.2)	562 (3.3)	563 (3.6)	517 (3.2)	516 (3.2)
Poland	2,394	2,568	4,962	9.9	533 (2.5)*	519 (2.7)	476 (2.4)	486 (2.5)*	502 (3.0)	508 (2.9)*
Portugal	1,957	2,034	3,991	10.0	548 (3.0)*	534 (2.8)	529 (4.1)	535 (3.4)	519 (4.6)	524 (3.8)
Romania	2,246	2,397	4,643	10.8	510 (4.8)*	495 (4.3)	481 (6.7)	484 (5.9)	505 (6.9)	506 (5.7)
Slovak Republic	2,736	2,825	5,561	10.4	540 (3.1)*	530 (2.8)	503 (4.0)	511 (3.9)*	528 (4.3)	536 (3.6)*
Slovenia	2,115	2,318	4,433	9.8	539 (2.2)*	523 (2.7)	508 (2.2)	518 (3.1)*	517 (2.8)	523 (3.4)
Spain	2,021	2,084	4,105	9.8	516 (2.5)	511 (2.8)	477 (3.1)	488 (3.4)*	500 (2.8)	510 (3.7)*
Sweden	2,193	2,289	4,482	10.7	549 (2.4)*	535 (2.5)	501 (2.5)	506 (2.4)	532 (3.0)	535 (3.2)
Total	36,592	38,276	74,868	10.3	540 (.6)*	528 (.6)	516 (.7)	522 (.7)*	518 (.8)	524 (.8)*

Note. Achievement test data after weighting to obtain nationally representative samples. Achievement test scores ranged from 5.0 to 870.2.
* Statistically significantly ($p < .05$) higher than the values of the other gender.

comparable on their age and on their amount of schooling. All countries applied a strict sampling procedure to allow analyses of nationally representative data. Sample selection was strictly monitored in every country to preserve high quality sampling standards (for more details regarding the sampling procedure, see Martin & Mullis, 2012).

Measures

Academic competencies. In TIMSS and PIRLS 2011, the students were administered academic competency measures assessing students' proficiency in reading (12 reading passages, 146 items), mathematics (180 items), and science (206 items).[1] The achievement tests had been developed on the basis of advice from substantive and statistical expert panels and the results of extensive pilot studies according to internationally aligned frameworks, ensuring, inter alia, measurement invariance across genders. Whereas the reading test items assessed reading literacy, both the mathematics and the science test items were constructed on the basis of a core curriculum that was comparable for all countries (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009).

The test items were administered in a multimatrix design in which 14 different booklets covering both mathematics and science and 13 booklets covering reading were randomly assigned to the students. Each student worked on two reading passages and approximately 24 to 30 mathematics and 24 to 30 science items. Although there was some time restriction, the TIMSS and PIRLS tests were primarily designed as power tests (Mullis, Martin, Kennedy, et al., 2009; Mullis, Martin, Ruddock, et al., 2009).

Gender equity. We used domain-specific gender equity indicators that would be theoretically expected to explain cross-national differences in academic achievement (Else-Quest & Grabe, 2012). These indicators were (a) women's access to education, as measured by the ratio of women and men with at least a secondary education level (typically at least 9 years of education completed) and by the ratio of the net tertiary school enrollment rates of women and of men (tertiary education programs comprised either theory-based programs to qualify for academic high-skill professions, with a duration of at least 3 years, or somewhat more practical programs to qualify for a direct entry into the labor force, with a duration of at least 2 years). Women's access to education "reflects the value of girls' education" in a society (Else-Quest & Grabe, 2012, p. 137). This value could lead girls and women to accordingly value their own academic development and consequently to engage in or to disengage from it. The second indicator was (b) women's participation in the labor market (as measured by the ratio of women's and men's participation in the labor market in the working age group between 15 and 64; participation regardless of how many hours worked). The last indicator of gender equity was (c) women's share of research positions (in percent of all research positions; head count, comprising both part-time and full-time employment). The two latter indicators might explain gender differences in academic achievement because working women, especially when working in research, might serve as role models for girls. This could make girls feel more self-confident about their abilities and their opportunities to participate later in society as an equal member of it, which could in turn promote their academic achievement (Else-Quest & Grabe, 2012; Else-Quest et al., 2010). Thereby, women's share of research positions, that is, in an academic top position, might be especially significant for the gender ratios at the top academic competency level.

The data for these indicators were taken from the United Nations Educational, Scientific, and Cultural Organization Institute for Statistics (http://data.uis.unesco.org); from the Organization for Economic Cooperation and Development (http://stats.oecd.org); and from the 2010 and 2011 Human Development Reports (United Nations Development Programme, 2010, 2011). Whenever possible, missing data for 2011 were replaced by the data from the preceding or the subsequent year (or the mean of both). For all measures, higher values indicate higher gender equity.

Analyses

Because of the multimatrix design, latent constructs were measured by plausible values. Generally speaking, plausible values result from applying missing value theory to estimates of latent construct values. The latent constructs in TIMSS and PIRLS are the reading, mathematics, and science competency values. These competency values are conceptualized as missing values. Therefore, multiple imputation strategies can be used to replace the missing values with reasonable estimates (Rubin, 1987). For this purpose, an imputation model including all variables that are part of the analysis model is needed to predict the most likely estimates.

For analyzing the relations across reading, mathematics, and science, a multidimensional IRT model was used as an imputation model (Martin & Mullis, 2012). As predictors, the students' scores on the test items as well as additional conditional variables were used. These conditional variables were principal components from a principal component analysis of all available student-level contextual data (e.g., family background, learning activities, domain-specific ability self-concept). To account for the uncertainty in the imputation strategy, five plausible values per student and domain were sampled and the variance of the statistic Q across the imputation values was added to the standard errors of Q (for a detailed description of the scaling procedure and the use of plausible values, see Martin & Mullis, 2012).[2] This scaling procedure preserved the correlational structure across the three subjects (multidimensional model). In addition, by using the conditional variables the measurement accuracy of the latent construct values increased, that is, the overall reliability of the three achievement scales was improved. The international median reliability was .82 for mathematics, .78 for science, and .88 for reading (Martin & Mullis, 2012). After executing the scaling procedure, each achievement scale was put on its own metric with an international mean of 500 and a standard deviation of 100.

To derive students' multidimensional proficiency profiles, we conducted a LPA (Gibson, 1966; Lazarsfeld & Henry, 1968). A

[1] Reading comprised reading intentions and reading comprehension. Mathematics comprised arithmetic, geometry, and handling data. Science comprised biology, physics/chemistry, and geography. Item examples can be obtained from http://timssandpirls.bc.edu/timss2011/international-released-items.html (TIMSS 2011) and http://timssandpirls.bc.edu/pirls2011/international-released-items.html (PIRLS 2011).

[2] Five plausible values were seen as optimal in terms of efficiency. More than five plausible values would have caused only marginally greater measurement accuracy (Schafer, 1999).

three-step approach was employed. In the first step, the total sample ($N = 74{,}868$) was separated into two to eight mixtures after we had weighted the data for the countries' different sample sizes so that every country contributed equally to the profile solution. The resulting seven mixture distributions were then compared based on the log-likelihood, the consistent Akaike information criterion, the Bayesian information criterion (BIC), and the sample-size-adjusted BIC (aBIC). After deciding for the mixture with the best data fit (Model 1; international model), the conditional means of the profiles were calculated and all students of the sample were assigned to the mixtures based on their most likely latent profile membership. Thus, the latent profiles were characterized by (a) their conditional means on the dimensions of the multidimensional distribution and (b) the percentage of students composing the profiles.[3]

In the second step, the conditional means were introduced as fixed parameters in national LPAs. That is, the country-specific multidimensional marginal distribution of students' achievement scores were separated into mixture distributions, in which both the number and the means of the mixtures were fixed at the values from the international model (Model 2; country-specific models with fixed means). Doing so allowed us to estimate the percentage of students forming the profiles and to assess how well the students from each national sample could be assigned to one of the profiles of the international model. The latter was indicated by both the relative entropy and the classification error rate.[4]

In the third step, Model 2 was applied again but, this time, the conditional means were free parameters; additionally, students' gender was included as a given manifest class (Model 3). Hence, the gender distribution within the national profiles could be examined. Due to the complex sampling procedure implemented in TIMSS and PIRLS, the students in the samples had had different probabilities of being selected. To adjust for these different selection probabilities and, thus, to make the sample representative of the desired population, we inversely weighted the units by their selection probabilities before applying Models 2 and 3.

Because five plausible values for the achievement domains were used, all analyses were performed five times (for each plausible value once). The results of these analyses were combined according to the formula by Rubin (1987). All analyses were conducted using Mplus 7.11 in combination with the full information maximum likelihood approach.

Results

Descriptive Statistics

Means and standard errors of the reading, mathematics, and science achievement test scores for the whole sample and for each country are displayed in Table 1 (see also Martin & Mullis, 2012; Mullis et al., 2012; Mullis, Martin, Foy, & Drucker, 2012). Although there was some variability across countries, girls consistently scored higher than boys in reading (average $d = 0.12$), and boys scored consistently but negligibly higher than girls in mathematics ($d = 0.06$). In science, there was also an average $d = 0.06$ in favor of boys, but the pattern of differences was rather inconsistent across countries.

Latent Profile Models

We first conducted an LPA using the samples of all 17 countries (weighted to be equal) to determine the international model (Model 1). Table 2 displays the fit criteria for the different possible international models. As can be seen, the consistent Akaike information criterion, the BIC, and the aBIC consistently decreased as the number of assumed profiles increased. Thus, the model with eight profiles achieved the best fit. However, as soon as the number of profiles exceeded seven, the number of students became extremely small for some profiles and therefore lacked substantive importance. In solutions with more than eight profiles, this pattern would have become even more extreme, given that the multidimensional distribution of the plausible values was approximately normal. Therefore, we chose the eight-profile solution as the upper bound of our analyses. When compared with the eight-profile model, the seven-profile model displayed somewhat more balanced proportions of students across the profiles, so that the numbers of students in the different profiles were of more practical importance (see Table 3, Columns 1 to 3). Because the difference in model fit between the eight-profile model and the seven-profile model was also very small and because the seven-profile model was the more parsimonious model and interpretable, we chose the seven-profile model as the international model.

In Hypothesis 1, we postulated that the international model would reveal student profiles that differ in their overall performance level across all three academic competencies rather than in the shape of their performance profile. In line with this prediction, the test scores of all achievement domains progressively and consistently increased from Profile 1 to Profile 7 (see Table 3, Columns 7, 9, and 11). No cross-nested structure for the profiles was observed. This suggests that the students could only be separated by different achievement levels on all three domains simultaneously and that no differentiation with respect to subject-specific strengths or weaknesses was possible.

In Hypothesis 2, we postulated that a reasonably large number of students in each country could reliably be assigned to one of the profiles of the international model built across all countries. We applied the international model in each country separately (with every country's data weighted to be representative), while holding the means constant (Model 2). The results indicated an acceptable assignment in 15 of the 17 countries (entropy $> .75$, classification error rate $< .25$). The assignment of the students from Romania and Malta was not as straightforward. More than a quarter of the students in these countries could not be assigned reasonably well into the international profile model. Nevertheless, a transfer from the international model to the country-specific distributions of students' achievement scores was possible without further restrictions. Thus, Hypothesis 2 was supported.

[3] We did not restrict the variances to be equal across the profiles.

[4] Relative entropy is the degree to which subjects can be clearly separated into the profiles. A value of 1 indicates a perfect fit, whereas a value of 0 indicates that the subjects could not be clearly separated into the profiles; classification error rate is the average posterior cross-classification probability. The posterior cross-classification probability of profile l is the likelihood that a student who is assigned to profile k will belong to profile l. Thus, the classification error rate represents the reliability of the profile solution. It ranges from 0 to 100% and is 0% when the profile solution is perfectly reliable.

Table 2
Comparison Between Models With Different Numbers of Latent Profiles (International Model; Model 1)

Number of profiles	Mean 2 × log-likelihood	CAIC	BIC	aBIC	Number of parameters
2	−1,266,986	2,533,992	2,534,084	2,534,052	10
3	−1,243,507	2,487,041	2,487,170	2,487,126	14
4	−1,231,360	2,462,756	2,462,756	2,462,865	18
5	−1,224,201	2,448,446	2,448,649	2,448,579	22
6	−1,220,405	2,440,862	2,441,102	2,441,020	26
7	−1,218,364	2,436,789	2,437,065	2,436,970	30
8	−1,217,217	2,434,503	2,434,816	2,434,708	34

Note. log-likelihood = the value of the corresponding fit-function when the estimation algorithm reaches the convergence criterion; CAIC = consistent Akaike information criterion; BIC = Bayesian information criterion; aBIC = sample-size-adjusted Bayesian information criterion. Smaller values indicate a better fit of the assumed mixture.

Gender Ratios in the Competency Profiles

To inspect the gender ratios (Research Question 3), we fixed the number of profiles and applied the model in every country, this time including gender in the analyses (Model 3). Overall, 73,331 of the 74,868 students (97.9%) could be assigned to one of the profiles. As can be seen in Table 3 (Column 6), boys were on average overrepresented at both the low and the high end and underrepresented in the middle range of the ability spectrum. Overall, boys outnumbered girls in Profile 1 by a ratio of 1.19 and in Profile 7 by a ratio of 1.24.

We then inspected variability in the gender ratios across the countries among the Profile 7 students (Research Question 4). Table 4 (Columns 2 to 4) shows that there was some considerable variation in the gender ratios. Although boys outnumbered girls in 14 of the 17 countries (even though the gender ratio of 1.15 in Lithuania was not statistically significant, $p > .01$), the gender ratios varied between 1.14 (Finland) and 1.67 (Czech Republic). Moreover, in two countries (Northern Ireland and Sweden), there was no difference between boys' and girls' percentage, and in one country (Ireland) there were slightly more girls than boys among the Profile 7 students (gender ratio: 0.96, *ns*).

To explain this variation (Hypothesis 5), we correlated the Profile 7 gender ratios with the values of the gender equity indicators (Table 4, Columns 5 to 8). The ratio of the net tertiary school enrollment rate (enrollment in postsecondary education programs with a duration of at least 2 or 3 years, qualifying for either direct labor force entry or academic high-skill professions) showed a small correlation with the gender ratios ($r = .19$).[5] This was however in the unexpected direction: In countries with a more equal tertiary enrollment rate, the gender ratios were more in favor of boys. The ratio of women to men with at least secondary education, women's labor market participation (regardless of hours worked) in the group of 15- to 64-year-olds, and women's share of research positions (head count, regardless of part- or full-time employment) correlated in the expected direction and in middle size with the gender ratios ($r = -.36$, $-.42$, and $-.33$, respectively). Thus, in countries where women are higher educated relative to men, show higher rates of labor market participation relative to men, and have a higher share of research positions, the gender ratios tended to be more balanced. Taken together in a multiple regression analysis, these indicators explained 28.7% of the gender ratio variance.

Discussion

In the present study, we investigated students' competency profiles and gender differences within these profiles. Furthermore, we inspected cross-national variability of the gender ratios among the top performers and tested whether this variability might be explained by societal gender equity. We achieved this by analyzing representative TIMSS and PIRLS 2011 data from 74,868 fourth graders from 17 European countries.

Interrelatedness of Academic Competencies

We derived an international profile model with seven profiles. As predicted in Hypothesis 1, the test scores in all achievement domains consistently increased from Profile 1 to Profile 7. Thus, the profiles unambiguously represented different ability levels simultaneously across different domains. This finding indicates that academic achievement in one domain is highly related to the achievement in the other two domains. This interrelatedness was so strong that possible relative domain-specific strengths and weaknesses as documented in other studies (e.g., Brunner et al., 2013; Wang et al., 2013) had no impact on the formation of the profiles when representative samples of students from a wide range of settings were investigated.

This gives a hint toward the importance of determinants shaping academic competencies across domains, such as the SES and the school environment. Professional intervention programs especially for low-SES children could be applied to foster their academic development. This could be accompanied by improvements of schools located in low-SES communities and of parent–school interaction. Chung (2015) and Kellaghan (2015) suggest a variety of opportunities to reach these goals, such as home support initiatives, early literacy and numeracy intervention programs, involving parents in school activities, or attracting highly qualified teachers to low-SES schools.

In line with Hypothesis 2, the profile patterns held across countries. Only in Malta and Romania were there slightly different profile patterns. Further analyses revealed that in these two countries, students with relative strengths and weaknesses appeared more often than in the other countries. Clarifying the reasons for the weaker interrelatedness of academic competencies in these countries might be an interesting purpose of future research.

Beyond the influence of the determinants already discussed, there might be further reasons for the interrelatedness we found. Different academic competencies might be needed when working on a specific competency test. For example, reading skills might be required for solving mathematics or science test items. Indeed, mathematics (and science) tasks often require students at least to some extent to read texts and to understand their meaning to grasp what is mathematically required (e.g., Abedi & Lord, 2001). Reading ability is also crucial in the process of gaining new knowledge in domains such as science or mathematics (Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999). In turn, knowledge acquired in different school domains might facilitate reading achievement because reading comprehension is derived from context information (Fukkink, 2005). This consideration has important

[5] Due to the small number of cases, we do not report p values but instead focus on the practical significance of the findings.

Table 3
Characteristics of the International Model (Model 1) Profiles: Number and Relative Frequency of Students (Overall and by Gender), Gender Ratios as Well as Means and Standard Errors of Students' Reading, Mathematics, and Science Scores

Profile	*n*	%			Gender ratio	Reading		Mathematics		Science	
		Overall	Girls	Boys		*M*	*SE*	*M*	*SE*	*M*	*SE*
1	783	1.0	1.0	1.1	1.19	251	27.5	273	31.3	243	26.2
2	2,993	4.0	3.7	4.4	1.13	345	23.5	358	20.8	343	23.7
3	8,514	11.4	11.2	11.7	1.01	419	16.9	419	13.6	417	15.8
4	18,193	24.3	24.6	23.7	.95	482	11.8	474	10.1	480	11.9
5	24,140	32.2	32.3	31.2	.97	538	9.0	527	9.1	538	9.2
6	16,681	22.3	22.3	22.5	1.03	592	7.3	580	7.8	594	7.4
7	3,564	4.8	4.9	5.4	1.24	655	7.0	644	8.2	657	7.2

Note. Gender ratio = number of boys for every girl; base-rate corrected average across countries after weighting to achieve nationally representative samples and after weighting countries' differences in sample size. Greater standard errors in Profiles 1 to 3 are primarily due to their smaller number of cases.

implications for fostering practices and instruction. If students are to be fostered in, for example, their mathematical ability, one should at the same time foster reading ability to achieve maximum learning results. It seems that high competency levels in several domains are a prerequisite for excellence in one particular domain. Therefore, instruction in a particular subject should not only focus on fostering the respective competency (e.g., fostering mathematical ability in mathematics instruction) but should be more holistic in the sense that different abilities should be focused on at the same time (e.g., fostering mathematical ability *and* reading ability in mathematics instruction).

One further question is whether the relations between the different domains might also be due to methodological factors, such as common method variance. However, if any, common method variance most often makes up an only negligible part of the overall variance (Jensen, 1998; Spector, 2006). This is the case even in questionnaire data which are actually thought of to be especially prone to factors causing common method variance such as social desirability, negative affectivity, or acquiescence (e.g., Spector, 2006). Thus, it seems unlikely that common method variance could explain a noticeable part of the interrelatedness between the different achievement domains that caused our LPA results.

Gender Differences in Academic Competencies

The differences in mean scores in all domains were small, supporting the gender similarities hypothesis (Hyde, 2005). However, we found that boys were overrepresented at both the low and the high end of the proficiency spectrum and underrepresented in the middle range. The more the competency profile departed from

Table 4
Gender Differences Among the Country-Specific Top Achieving Students (Profile 7; Model 3) as Well as Gender Equity Indicators

Country	Students in Profile 7		Gender ratio	Women's access to education		Ratio of women to men participating in the labor market	Women's share of research positions (%; head count)
	% of girls (*SE*)	% of boys (*SE*)		Ratio of women to men with secondary or higher education	Ratio of women to men enrolled in tertiary education		
Austria	5.3 (.5)	6.9 (.5)	1.30*	.78	1.18	.87	29.0
Czech Republic	2.4 (.3)	4.0 (.4)	1.67*	.98	1.42	.79	28.2
Finland	4.2 (.4)	4.8 (.4)	1.14*	1.00	1.23	.94	32.1
Germany	5.0 (.5)	6.5 (.6)	1.30*	.98	.93	.87	26.7
Hungary	4.3 (.4)	5.1 (.4)	1.19*	.96	1.32	.83	31.7
Ireland	5.5 (.5)	5.3 (.5)	.96	1.01	1.04	.81	32.4
Italy	2.9 (.4)	4.1 (.4)	1.41*	.86	1.43	.71	34.9
Lithuania	2.6 (.3)	3.0 (.3)	1.15	.96	1.47	.91	52.1
Malta	3.7 (.5)	5.0 (.5)	1.35*	.88	1.34	.53	26.9
Northern Ireland[a]	3.0 (.4)	3.0 (.4)	1.00	1.01	1.36	.85	37.7
Poland	2.9 (.3)	4.0 (.4)	1.38*	.95	1.55	.81	38.6
Portugal	3.8 (.4)	4.4 (.5)	1.16*	.96	1.19	.89	46.4
Romania	4.1 (.4)	4.7 (.4)	1.15*	.93	n/a	.78	46.1
Slovak Republic	4.4 (.4)	5.7 (.4)	1.30*	.93	1.55	.79	42.6
Slovenia	3.8 (.4)	4.8 (.4)	1.26*	.74	1.70	.90	36.4
Spain	4.0 (.4)	5.3 (.5)	1.33*	.94	1.23	.84	38.7
Sweden	5.2 (.5)	5.2 (.5)	1.00	1.01	1.51	.94	37.2

Note. Gender ratio = number of boys for every girl; n/a = not available.
[a] Gender equity data from the United Kingdom.
* Significantly ($p < .01$) different from 1.00.

the middle range, the more boys were represented. Thus, even when reading, mathematics, and science are considered simultaneously, boys are still overrepresented at the high end of the competency distribution. Besides gender differences in interests and career choices, this finding might contribute to an explanation for why there are more men than women among the top performers in societal key positions (Ceci et al., 2009).

We also found boys to be overrepresented among the weakest performing students. Against the background of boys being less successful than girls in the educational systems in many countries around the world (e.g., Spinath, Eckert, & Steinmayr, 2014; Voyer & Voyer, 2014), this is an important finding, too. Single studies have demonstrated that boys' lower school achievement is likely to be explained by personality and motivational factors (e.g., Duckworth & Seligman, 2006; Kessels & Steinmayr, 2013; Steinmayr & Spinath, 2008). As it is possible that factors at other levels also contribute to the underperformance of boys in school, further studies should investigate whether sociocultural factors explain the overrepresentation of boys at the low end of the ability distribution.

What causes the greater variability in competencies in boys compared to girls? One explanation might be that boys do not only differ more in competencies but already in other characteristics influencing those competencies. Whereas boys and girls do not systematically differ in their variance in different personality and motivational determinants of academic success (Steinmayr & Spinath, 2008), they differ in their variance in general mental ability (GMA). Studies using large representative samples have found that there are no or only negligible gender differences in mean GMA, but that males display greater variability than females, which then results in a male overrepresentation both in the upper and in the lower tail of the GMA distribution (e.g., Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Johnson, Carothers, & Deary, 2008; Strand, Deary, & Smith, 2006). However, the gender ratio in the upper tail of the GMA distribution has been narrowed within the last decades (see Ceci et al., 2009; Lohman & Lakin, 2009). Thus, sociocultural factors, among others (Ceci et al., 2009; Hyde, 2014), are likely to contribute to the gender gap in the upper tail of the GMA distribution.

Likewise, we found clear hints that gender differences in the highest academic competency profile are to a substantial degree due to sociocultural factors. When we inspected the gender ratios among the top academic performers, one striking finding was that there was some considerable variation in these gender ratios across the countries. Moreover, we found evidence for a significant role of society's gender equity in both education and professional life. Gender ratios in favor of boys were smaller in countries where women had more secondary or higher educational levels, were more present in the labor market, and held more research positions. It would be interesting to investigate whether these sociocultural factors also contribute to gender differences in the upper tail of the GMA distribution.

The correlation between the tertiary school enrollment ratio and the gender ratios was however in an unexpected direction. This might be explained by a selection effect. In countries where female enrollment rates are lower than male enrollment rates, the girls and women who are enrolled are a more strictly ability-selected group than are boys or men in these countries. This could lead to more balanced gender ratios (when corrected for the base rate of females and males attending school). This interpretation is supported by studies showing that gender gaps especially among high mathematics and science achievers were more balanced in countries with higher power distance (i.e., with high segregation of social groups and a high tolerance toward inequity; see Reilly, 2012). Females in those countries might not only compose a relatively highly abled group but might also be more motivated than males to gain education to overcome their lower social status (Reilly, 2012).

Of course, correlations do not allow for causal conclusions. The causation might also work in the other direction than in the direction discussed above. For example, the fact that there are more women participating in the labor market and holding more research positions might also be a consequence—and not a cause—of smaller gender differences in mean or high levels of academic competencies (see, e.g., Ceci et al., 2009). However, as Else-Quest et al. (2010) already noted, the hypothesis of such causation cannot explain why gender differences in academic competencies occur in some countries but not in others. Other possibilities might be a reciprocal causation (see also Else-Quest et al., 2010) or the influence of a superordinate factor such as gender stereotypes (Miller, Eagly, & Linn, 2015). Answering the question of the causal direction would be an interesting and challenging issue for future research. In any case, the cross-national variability of the gender differences demonstrates that for the genesis of gender differences among the top performers, sociocultural factors definitely play an important role.

Limitations and Future Directions

In our study, we presented findings from large nationally representative samples of fourth graders from 17 European countries sharing the same supranational educational policy initiative. Despite the advantages of such a sample selection, it might be desirable for future studies to include as many culturally different countries as possible to evaluate the cross-national variability of the gender ratios at the high end of academic competencies even more comprehensively. Furthermore, societal gender equity could not explain the entire gender ratio variance. There must be additional factors at work causing the gender ratios. To unravel them, it could be useful to investigate more thoroughly which part of the gender ratios is due to differences in mean and which part is due to differences in variability, because the causes for differences in mean might be different from the causes for differences in variability (Humphreys, 1988).

As a final limitation, we studied achievement test scores only in reading, mathematics, and science. Although these three competencies are regarded as the three core academic competencies with the most predictive power of important life outcomes, it would have been desirable to include even more competencies taught at school.

Conclusions

We showed that (a) elementary school students across 17 European countries could (only) be clustered according to their achievement level across all three domains readings, mathematics, and science, and not according to their performance profile shape. We also showed that (b) boys were more likely than girls to perform at the top level on academic performance tests. However, we found that (c) there was some cross-national variability in this tendency and that societal gender equity partly explained this variability. This speaks to an important role of sociocultural factors for the explanation of gender differences among the top academic performers.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 24,* 219–234. http://dx.doi.org/10.1207/S15324818AME1403_2

Bos, W., Wendt, H., Ünlü, A., Valtin, R., Euen, B., Kasper, D., & Tarelli, I. (2012). Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland [Achievement profiles of fourth-graders in Germany]. In W. Bos, H. Wendt, O. Köller, & C. Selter (Eds.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 269–301). Münster, Germany: Waxmann.

Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Krauss, S., & Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence, 41,* 378–395. http://dx.doi.org/10.1016/j.intell.2013.05.009

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence, 36,* 403–421. http://dx.doi.org/10.1016/j.intell.2007.11.002

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity-propensity analysis. *Contemporary Educational Psychology, 32,* 599–629. http://dx.doi.org/10.1016/j.cedpsych.2006.09.002

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135,* 218–261. http://dx.doi.org/10.1037/a0014412

Chung, K. K. H. (2015). Socioeconomic status and academic achievement. In J. D. Wright (Series Ed.) & C. Byrne, P. Schmidt, & C. McBride-Chang (Vol. Eds.), *International encyclopedia of the social and behavioral sciences: Education* (pp. 924–930). Philadelphia, PA: Elsevier. http://dx.doi.org/10.1016/B978-0-08-097086-8.92141-X

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35,* 13–21. http://dx.doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey. *Intelligence, 31,* 533–542. http://dx.doi.org/10.1016/S0160-2896(03)00053-9

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98,* 198–208. http://dx.doi.org/10.1037/0022-0663.98.1.198

Else-Quest, N. M., & Grabe, S. (2012). The political is personal: Measurement and application of nation-level indicators of gender equity in psychological research. *Psychology of Women Quarterly, 36,* 131–144. http://dx.doi.org/10.1177/0361684312441592

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103–127. http://dx.doi.org/10.1037/a0018053

European Commission. (2015, August 24). Strategic framework: Education and training 2020. Retrieved from http://ec.europa.eu/education/policy/strategic-framework/index_en.htm

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. http://dx.doi.org/10.3102/00346543062001061

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *Journal of Personality and Social Psychology, 97,* 517–532. http://dx.doi.org/10.1037/a0016030

Fukkink, R. G. (2005). Deriving word meaning from written context: A process analysis. *Learning and Instruction, 15,* 23–43. http://dx.doi.org/10.1016/j.learninstruc.2004.12.002

Gibson, W. A. (1966). Latent structure analysis and test theory. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 78–88). Chicago, IL: Science Research Associates.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Diversity: Culture, gender, and math. *Science, 320,* 1164–1165. http://dx.doi.org/10.1126/science.1154094

Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63,* 94–105. http://dx.doi.org/10.3102/00346543063001094

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269,* 41–45. http://dx.doi.org/10.1126/science.7604277

Helwig, R., Rozek-tedesco, M. A., Tindal, G., Heath, B., & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93,* 113–125. http://dx.doi.org/10.1080/00220679909597635

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58,* 47–77. http://dx.doi.org/10.3102/00346543058001047

Humphreys, L. G. (1988). Sex differences in variability may be more important than sex differences in means. *Behavioral and Brain Sciences, 11,* 195–196. http://dx.doi.org/10.1017/S0140525X00049402

Hunt, J. (2016). Why do women leave science and engineering? *ILR Review, 69,* 199–226. http://dx.doi.org/10.1177/0019793915594597

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373–398. http://dx.doi.org/10.1146/annurev-psych-010213-115057

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence. A new look at the old question. *Perspectives on Psychological Science, 3,* 518–531. http://dx.doi.org/10.1111/j.1745-6924.2008.00096.x

Kellaghan, T. (2015). Family and schooling. In J. D. Wright (Series Ed.) & C. Byrne, P. Schmidt, & C. McBride-Chang (Vol. Eds.), *International encyclopedia of the social and behavioral sciences: Education* (pp. 751–757). Philadelphia, PA: Elsevier. http://dx.doi.org/10.1016/B978-0-08-097086-8.92005-1

Kessels, U., & Steinmayr, R. (2013). Macho-man in school: Toward the role of gender role self-concepts and help seeking in school performance. *Learning and Individual Differences, 23,* 234–240. http://dx.doi.org/10.1016/j.lindif.2012.09.013

Kuncel, N. R., Ones, D. S., & Hezlett, S. A. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181. http://dx.doi.org/10.1037/0033-2909.127.1.162

Kurtz-Costes, B. (2015). Families as educational settings. In J. D. Wright (Series Ed.) & C. Byrne, P. Schmidt, & C. McBride-Chang (Vol. Eds.), *International encyclopedia of the social and behavioral sciences: Education* (pp. 731–737). Philadelphia, PA: Elsevier. http://dx.doi.org/10.1016/B978-0-08-097086-8.92004-X

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136,* 1123–1135. http://dx.doi.org/10.1037/a0021276

Lohman, D. F., & Lakin, J. M. (2009). Consistencies in sex differences on the Cognitive Abilities Test across countries, grades, test forms, and cohorts. *The British Journal of Educational Psychology, 79,* 389–407. http://dx.doi.org/10.1348/000709908X354609

Machin, S., & Pekkarinen, T. (2008). Assessment: Global sex differences in test score variability. *Science, 322,* 1331–1332. http://dx.doi.org/10.1126/science.1162573

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96,* 56–67. http://dx.doi.org/10.1037/0022-0663.96.1.56

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., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107,* 631–644. http://dx.doi.org/10.1037/edu0000005

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

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

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Nowell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39,* 21–43. http://dx.doi.org/10.1023/A:1018873615316

Organization for Economic Cooperation and Development (OECD). (2014). *PISA 2012 results: What students know and can do: Student performance in mathematics, reading and science* (Rev. ed., Vol. 1). OECD Publishing, Paris.

Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology, 114,* S138–S170. http://dx.doi.org/10.1086/589252

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE, 7,* e39904. http://dx.doi.org/10.1371/journal.pone.0039904

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107,* 645–662. http://dx.doi.org/10.1037/edu0000012

Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence, 38,* 231–241. http://dx.doi.org/10.1016/j.intell.2010.01.002

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

Rindermann, H. (2007). The *g* factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-Tests across nations. *European Journal of Personality, 21,* 667–706. http://dx.doi.org/10.1002/per.634

Rost, D. H. (2013). *Handbuch Intelligenz* [Handbook of intelligence]. Weinheim, Germany: Beltz.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley-Interscience. http://dx.doi.org/10.1002/9780470316696

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15. http://dx.doi.org/10.1191/096228099671525676

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

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9,* 221–232. http://dx.doi.org/10.1177/1094428105284955

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research, 56,* 230–243. http://dx.doi.org/10.1080/00131881.2014.898917

Steinmayr, R., Dinger, F. C., & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. *European Journal of Personality, 24,* 535–550. http://dx.doi.org/10.1002/per.755

Steinmayr, R., Dinger, F. C., & Spinath, B. (2012). Motivation as a mediator of social disparities in academic achievement. *European Journal of Personality, 26,* 335–349. http://dx.doi.org/10.1002/per.842

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality, 22,* 185–209. http://dx.doi.org/10.1002/per.676

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19,* 80–90. http://dx.doi.org/10.1016/j.lindif.2008.05.004

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A U.K. national picture. *The British Journal of Educational Psychology, 76,* 463–480. http://dx.doi.org/10.1348/000709905X50906

United Nations Development Programme. (2010). *Human development report 2010*. Houndmills, United Kingdom: Palgrave Macmillan.

United Nations Development Programme. (2011). *Human development report 2011*. Houndmills, United Kingdom: Palgrave Macmillan.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174–1204. http://dx.doi.org/10.1037/a0036620

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101,* 817–835. http://dx.doi.org/10.1037/a0016127

Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24,* 770–775. http://dx.doi.org/10.1177/0956797612458937

Received September 10, 2015
Revision received June 2, 2016
Accepted June 3, 2016 ■