

Special section  
MATHEMATICAL LEARNING:  
NEW PERSPECTIVES AND CHALLENGES

**Sex differences in mathematical achievement:  
Grades, national test, and self-confidence**

Marina S. Egorova\*, Yulia D. Chertkova

*Faculty of Psychology, Lomonosov Moscow State University, Moscow, Russia*

\*Corresponding author. E-mail: ms\_egorova@mail.ru

Academic achievement, which is inherently an indicator of progress in the curriculum, can also be viewed as an indirect measure of cognitive development, social adaptation, and motivational climate characteristics. In addition to its direct application, academic achievement is used as a mediating factor in the study of various phenomena, from the etiology of learning disabilities to social inequality. Analysis of sex differences in mathematical achievement is considered particularly important for exploring academic achievement, since creating an adequate educational environment with equal opportunities for boys and girls serves as a prerequisite for improving the overall mathematical and technical literacy that is crucial for modern society, creates balanced professional opportunities, and destroys traditional stereotypes about the roles of men and women in society.

The objective of our research was to analyze sex differences in mathematical achievement among high school students and to compare various methods for diagnosing academic performance, such as school grades, test scores, and self-concept.

The results were obtained through two population studies whose samples are representative of the Russian population in the relevant age group. Study 1 looked at sex differences in math grades among twins ( $n = 1,234$  pairs) and singletons ( $n = 2,227$ ) attending high school. The sample of Study 2 comprised all twins who took the Unified State Examination in 2010–2012. The research analyzed sex differences in USE math scores across the entire sample and within the extreme subgroups. It also explored differences between boys and girls in opposite-sex dizygotic (DZ) twin pairs.

The key results were as follows. No difference in mathematical achievement was observed between twins and singletons. Sex differences were found in all measures of mathematical achievement. Girls had higher school grades in math than boys, while boys outperformed girls in USE math scores. Boys were more variable and there were more

boys at the right tail of the distribution. Girls with a positive math self-concept did better than boys on math tests. In groups of opposite-sex DZ twins, differences between the USE math scores of girls and boys were not significant.

The results obtained are presumed to correspond more closely to assumptions about the roles of non-cognitive factors of variation in mathematical ability than the mathematical ability theory.

**Keywords:** mathematical achievement, sex differences, school grades, math tests, self-concept

## **Introduction**

Despite the multitude of approaches to analyzing academic achievement in mathematics, some topics are far in the lead based on the number of publications, the intensity of discussion, and the variety of proposed theoretical models. These topics include the nature of sex differences in mathematical achievement—their size, change over time, causes and consequences for society. Despite extensive research into sex differences in mathematical achievement, many controversial issues and contradictions remain.

The objective of our research was to compare various methods for assessing mathematical achievement and to analyze sex differences observed with the use of the various methods.

### ***Measures of mathematical achievement***

Three types of measures are generally used to assess academic performance overall and mathematical achievement in particular: (a) school grades in individual subjects or, more frequently, grade point average (GPA) (for example, Kimball, 1989; McClure et al., 2011; Voyer & Voyer, 2014); (b) results of cognitive ability tests, standardized national assessments (such as the Graduate Record Examination in the U.S., the National Curriculum Tests in the UK, and the Unified State Examination, or USE, in Russia), and international tests that measure literacy and competency (such as the Program for International Student Assessment, or PISA) (for example, Benbow & Stanley, 1980, 1982; Hyde & Linn, 2006; Strand et al., 2006; Lohman & Lakin, 2009; Lindberg et al., 2010); and (c) self-assessment of mathematical achievement, which is frequently used in recent studies instead of direct assessment of academic achievement (for example, Spinath et al., 2008; Chamorro-Premuzic et al., 2010; Luo et al., 2011; Marsh et al., 2015; Seaton et al., 2014). Self-assessment is associated with a wide range of indicators, including direct self-appraisal of academic performance, achievement attitudes (self-concept, self-confidence, self-efficacy), extrinsic and intrinsic motivation, school anxiety, and personal indicators linked with subjects' own assessment of what they do better or worse—for example, selection of disciplines for advanced study, choice of college major, preference for a particular profession, etc.

All of these measures are associated with particular aspects of academic success and, predictably, correlate with each other. This creates the illusion that the measures are interchangeable in research. However, although the correlation in various measures of academic achievement is generally significant, it is not always very

high and, more importantly, the significant role that self-assessment of mathematical ability plays in mathematical achievement is not sufficient grounds for viewing it as a direct indicator of mathematical achievement.

Average correlation between math grades and test scores does not exceed 0.5 and is frequently even lower (Cucina et al., 2016). Thus, correlation between math grades and math test scores in studies with representative samples was in the 0.35–0.37 range for seventh-graders (Marsh et al., 2005) and in the 0.27–0.39 range for ninth-graders (Möller et al., 2014). The correlation of math self-concept with math grades and with test scores was similar at 0.38–0.44 and 0.28–0.32 (Marsh et al., 2005).

Differences among mathematical achievement indicators (grades, test scores, and self-concept) are also evident from analysis of their correspondence with other psychological traits.

Intelligence correlates more strongly with test scores than with school grades. For example, a comparison of intelligence measured by the Berlin Intelligence Structure (BIS) and the results of the Trends in International Mathematics and Science Study (TIMSS) yielded a correlation of 0.51 (Hofer et al., 2012). At the same time, a meta-analysis of the link between school grades in math and the mathematical subtests of various IQ tests found an average correlation of 0.43 based on 14 studies. The general factor in intelligence (*g*), diagnosed using the U.S. Armed Services Vocational Aptitude Battery (ASVAB), is linked with scholastic performance at about the same level: Correlation with GPA was 0.44, while correlation with the math subtests of the ASVAB — Arithmetic Reasoning, Numerical Operations, and Mathematics Knowledge — was 0.39, 0.18, and 0.42, respectively (Roth et al., 2015).

Self-control and procrastination correlate more closely with self-reported school grades (0.52, -0.32) than with test scores (0.33, -0.04) (Hofer et al., 2012). Self-regulation has strong links with math grades and no correlation with math test scores (Morosanova et al., 2014). There also exist differences based on personality traits: College students pursuing degrees in mathematical sciences exhibited lower neuroticism scores than those studying humanities (Vedel, 2016).

Differences between school grades and test scores depend significantly on the distinctive roles these play in the formation of math self-concept. Math grades earned through direct interaction with a teacher are believed to have more impact than test scores, which can be viewed as formal indicators that are not related to goals set by students in their studies (Trautwein et al., 2006; Simzar et al., 2015). The link between school grades and traditional definitions of success makes grades a more effective incentive for the formation of a positive or negative math self-concept than standardized test scores (Skaalvik & Skaalvik, 2002).

A number of theoretical models address the correlation between academic achievement and self-concept. An analysis of the cause-and-effect relationship between self-concept and academic achievement based on school grades and test scores included a thorough review of the three models, covering three possible options for correlation with academic achievement indicators: the self-enhancement model, which presumes the influence of motivational components of self-concept on academic achievement (Marsh & Yeung, 1997); the skill development model, which focuses on the importance of academic achievement for the formation of

self-concept (Byrne, 1996); and the reciprocal effects model, which views academic achievement as a precursor to self-concept and self-concept as the basis for the formation of academic achievement (Marsh, 1990; Marsh et al., 1999; Valentine et al., 2004; Marsh & Craven, 2006).

A reciprocal internal/external frame of reference model has been formed on the basis of these three models (Marsh & Köller, 2004; Marsh et al., 2015). The first steps towards the establishment of this model were made quite a while ago (Marsh, 1986). According to this model, a student's self-concept of his or her performance in a particular school subject is formed based on both external and internal comparisons (in other words, with two frames of reference): first, comparing one's own achievement with that of other students, and, second, comparing one's achievement in various academic domains. In the former (social comparison), academic achievement is a determinant of self-concept (high math grades compared to other students improve one's math self-concept). In the latter (ipsative comparison), academic achievement in one domain lessens perceived ability in other domains (if one's math grades are higher than one's literature grades, subsequent successes in math will reduce one's relative literature self-concept, regardless of how one's literature grades compare to those of classmates).

The reciprocal internal/external frame of reference model has been supported by a number of experimental studies (for example, Möller et al., 2011; Xu et al., 2013; Möller et al., 2014; Niepel et al., 2014; Marsh et al., 2015). A meta-analysis of the results of 69 studies with a cumulative sample of 125,308 individuals clearly demonstrated the characteristics of different academic achievement indicators (Möller et al., 2009). The meta-analysis focused on studies where academic achievement was diagnosed based on school grades, test scores, and self-reports (including affective, motivational, and cognitive components, as well as self-efficacy). The key results of the meta-analysis were as follows:

1. Average correlation between mathematical achievement and math self-concept for the entire sample was 0.43. Correlation between verbal achievement and verbal self-concept was slightly lower (0.35).

Correlation between mathematical achievement and verbal achievement was higher for test scores than for school grades (0.74 vs. 0.54), while correlation between mathematical and verbal self-concepts was lower for test scores than for school grades (0.37 vs. 0.50).

2. Correlation between mathematical achievement and verbal achievement for all indicators (without separation between school grades and test scores) was significantly higher than the correlation between math and verbal self-concepts (0.67 vs. 0.10).

3. Analysis showed that paths leading from mathematical achievement and verbal achievement to corresponding self-concepts were positive (0.61 and 0.49), while cross-paths (mathematical achievement–verbal self-concept and verbal achievement–math self-concept) were negative (-0.21 and -0.27).

These results support the assumptions of the reciprocal internal/external frame of reference model and—of special importance in the context of the present article (which seeks in particular to compare various measures of mathematical achievement)—demonstrate the inadequacy of viewing different indicators of mathematical achievement as interchangeable.

### ***Sex differences in mathematical achievement measures***

Assessments of sex differences in mathematical achievement vary regardless of how they are measured—teacher-assigned school grades, test scores, or self-reporting. There are some contradictions in the data for all cases: Some studies show that boys do better, others that girls perform more strongly, and still others demonstrate a lack of sex differences. Nevertheless, the key trends are clear.

The first comprehensive review of sex differences in math grades (Kimball, 1989) established that girls get better math grades and that their superior performance (higher grades) can be seen as early as elementary school. This has been supported by later reviews (for example, Amrein & Berliner, 2002; Ding et al., 2006). In middle school, girls generally have a small advantage, which increases in high school and declines somewhat in college and beyond.

The first meta-analysis of school grades was conducted using 369 samples of school and college students from various countries (Voyer & Voyer, 2014). It confirmed the aforementioned age dynamics of sex differences. In particular, girls were furthest ahead of boys during adolescence: in 14 studies of high school students  $d = -0.18$  (hereinafter, a negative Cohen's  $d$  indicates that girls outperform boys). However, average differences across the samples—while showing better performance by girls—were minuscule ( $-d = 0.07$ ). Subsequent analysis of the results of studies with small samples that were not included in the meta-analysis also demonstrated a certain, but quite modest, advantage for girls ( $-d = 0.11$ ).

Based on the results of standardized tests, the mathematical achievement of boys and girls is completely reversed: Boys have higher scores based on standardized methods of accessing mathematical achievement. Boys do better on mathematical subtests of cognitive tests (Strand et al., 2006; Lohman, Lakin, 2009) and outperform girls on national math tests (Hyde et al., 1990; Else-Quest, 2010; Lindberg et al., 2010) as well as international mathematics competency tests (Mullis et al., 2008; Nosek et al., 2009). There are also more boys in the highest-scoring groups (Benbow & Stanley, 1980; Benbow, C. P., 1988; Wai et al., 2010; Korpershoek et al., 2011).

Over the past 40 years, the gap between boys and girls in math test scores has consistently decreased, but it has not completely closed. Instead, it has plateaued at the same level (Ceci et al., 2014; Reilly et al., 2015; Wang, Degol, 2016). For example the average effect size ( $d$ ) for national performance data in National Assessment of Educational Progress mathematics (almost 2 million students) is 0.10 (i.e., sex differences are minimal).

Studies of math self-concept generally demonstrate higher self-confidence and self-efficacy for boys (Kling et al., 1999; Syzmanowicz & Furnham, 2011; Novikova & Kornilova, 2012). A meta-analysis of 54 studies conducted between 1997 and 2009 showed that self-reports of mathematical/logical intelligence by boys were much higher: almost half of a standard deviation above self-reports by girls. Only one study in the meta-analysis displayed higher self-reports by girls. Four others showed minuscule differences ( $d = 0.06$ ). For the entire sample, effect size ( $d$ ) reached 0.44 (Syzmanowicz & Furnham, 2011).

Thus, research results indicate that girls have higher mathematical achievement based on school grades and lower mathematical achievement based on test scores and math self-concept.

The research described below analyzed the manifestation of sex differences in mathematical achievement assessed using different measures. The new insights offered by the research relate first of all to the juxtaposition between the sex differences in mathematical ability observed in the general population and the sex differences observed in groups that self-select for STEM fields, and second to the analysis of within-family sex differences in mathematical ability.

## **Method**

Results from two studies were used to analyze sex differences in mathematical ability among high school students. The objective of the studies was to compare academic achievement of twins and singletons, but since they were carried out with samples representative of the general population of the relevant age, their results also provide a good illustration of sex differences in academic achievement.

### *Study 1*

The objective of the first study was to compare scholastic achievement of twins and singletons. The sample included monozygotic (MZ) twins, single-sex DZ twins, and opposite-sex DZ twins (total of 2,282 pairs), as well as singletons (4,065) from the same grades as the twins. The age of twins and singletons ranged from 8 to 17 (grades 2–11). The sample included about 2% of all school-age twins residing in Russia at the time of the study and was representative of the Russian school-age twin population based on socioeconomic status (SES), structure of family of origin, and characteristics of the region of residence—i.e., population size (from 3,000 to 10 million), economic development, and geographic location.

The present article only addresses mathematical achievement of high school students (grades 8–11). The sample included 1,234 pairs of twins (2,468 individuals: 1,315 girls and 1,153 boys, i.e., 53.28% vs. 46.72%) and 2,227 singletons (1,124 girls and 1,103 boys, i.e., 50.47% vs. 49.53%). The age of study participants ranged from 12 to 17 ( $M = 15.0$ ,  $SD = 1.43$ ).

Indicators: final school grades (2 as the lowest through 5 as the highest) in two academic subjects (algebra and geometry).

### *Study 2*

The objective of the second study was to compare scholastic achievement of twins and singletons based on the results of the Unified State Examination, which is taken by all high school graduates in Russia. Two academic subjects are mandatory components of the exam: Russian language and math. Unless excused for health reasons, all students must take the USE in order to get a high school diploma. Students applying to institutions of higher education must take additional exams in subjects based on their fields of study.

It should be noted that in the case of population studies, such as our research, the factor that the USE is most criticized for (violations of testing procedures) is not relevant for comparing mean group scores (e.g., twins vs. singletons or boys vs. girls), because the error is the same for all groups.

The sample of Study 2 includes all twins residing in Russia who took the USE in 2010–2012. Twin pairs were selected using the following algorithm: Students were



classified as twins if they shared a last name, patronymic, and date of birth, and took the test at the same location.

For the purposes of this study (comparing sex differences in school grades vs. in USE scores), all twins over the age of 19 were excluded from the sample. As a result, data were obtained on the USE test scores of 22,320 twins (11,160 single-sex and opposite-sex twin pairs, including 12,760 girls and 9,560 boys, i.e., 57.17% vs. 42.83%). The age of study participants ranged from 14 to 19 ( $M = 16.5$ ,  $SD = 0.59$ ).

In addition, a subsample of students from the entire sample of Study 2 was reviewed: individuals who took an optional physics exam in addition to the mandatory math exam (i.e., are likely to apply to institutions that specialize in STEM fields). From this viewpoint, a decision to take an optional physics exam is a good indicator of a positive math self-concept and self-confidence in mathematics.

The subsample of students who took the USE physics exam was 5,870 (1,705 girls and 4,165 boys, i.e., 29.05% vs. 70.95%).

Indicators: USE scores in mathematics, USE scores in mathematics of those who take an optional physics exam. USE scores range from 0 to 100.

The results of the two studies make it possible to analyze sex differences based on three indicators of mathematical achievement: math grades assigned by the teacher at the end of the school year, USE scores, and mathematical self-confidence (as assessed based on the selection of a specialization related to advanced study of mathematics).

Data processing was carried out using R-3.2.3 software (Wooden Christmas Tree). Student's  $t$ -distribution, the Pearson  $\chi^2$  criterion, and Cohen's effect size  $d$  were used to assess the significance of inter-group differences. Cohen's effect size  $d$  was calculated as the difference between the mean scores of boys and girls, divided by the average standard deviation for the groups of boys and girls. A positive  $d$  indicates that boys outperformed girls, while a negative  $d$  means that girls scored higher. The further  $d$  is from 0, the greater the sex difference in the given characteristic.

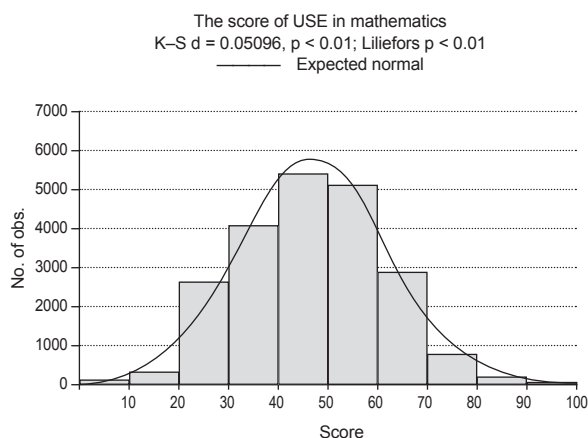
## Results

### *Frequency distribution of school grades and USE math scores*

Less than 0.5% of high school students received a final grade of 2 (on a scale of 2–5). It is unlikely that this reflects the number of students who successfully completed the curriculum. More plausibly, this indicator demonstrates tacit grading practices. A grade of 2 is an extraordinary event that has severe consequences for both students and teachers. Students who receive 2s may have problems advancing to the next grade level: The school has to schedule an additional exam in the fall and students who do not pass this exam must repeat the year. Students who receive 2s in their graduating year do not get their high school diploma. 2s also have consequences for teachers, because they are interpreted as an indicator of ineffective teaching and insufficient attention to weaker students. As a result, teachers tend to avoid “making trouble” for the students and for themselves, and “inflate” final grades of 2 to 3s.

Almost half of the students in grades 8–11 received 3s in math; 40% earned 4s and 12.5% got 5s. Thus, final grades highlight students with strong ability and interest in mathematics, but do not differentiate students who performed poorly from those who completely failed to progress through the curriculum.

Unlike with school grades, the distribution of USE math scores is closer to a standard bell curve (Figure 1). Scores in our sample range from 0 to 100 ( $M = 47.13$ ,  $SD = 15.07$ ).



**Figure 1.** USE Math score (frequency distribution)

USE scores are not only more useful for differentiating students based on mathematical ability, but also have a higher validity. Indirect proof of this is the balance in academic achievement of urban vs. rural dwellers.

It has been shown many times that the higher average SES of city residents, better access to high-quality education, a comprehensive extracurricular education offering, and a stronger orientation towards education all result in better academic performance by students who live in cities. Our study supports this: Students from urban communities scored much higher on the USE than students from rural areas (47.91 vs. 45.31,  $p < 0.001$ ).

The balance in the final school grades of urban vs. rural students is quite different. There is no difference in the mathematical achievement of urban and rural students (3.65 vs. 3.67,  $p = 0.53$ ). It is likely that teachers assign grades through a comparative assessment of students in their class, rather than based on an abstract federal education standard: Grades reflect relative performance (children who learned more than their classmates got the highest grades) rather than an absolute criterion (how well the curriculum had been absorbed).

Thus, even though the USE has been strongly criticized as a means of final testing, our data show that USE scores are more meaningful for comparative assessment of mathematical achievement than school grades.

### ***Comparing mathematical achievement of twins vs. singletons***

Assessing differences between twins and singletons is necessary in order to determine whether conclusions drawn from twin studies can be rightfully extended to



the overall population, which is mostly made up of singletons (95.2% in our sample). Therefore, the first objective of the research was to compare the mathematical achievement of twins and singletons in grades 8–11.

The frequency distribution of final grades in algebra and geometry (Table 1) does not show a significant difference between twin and singleton samples ( $\chi^2 = 0.26$ ,  $p = 0.88$  for algebra;  $\chi^2 = 0.23$ ,  $p = 0.89$  for geometry).

**Table 1.** Distribution of final grades in algebra and geometry (percent of sample) for students in grades 8–11

Final Grade	Twins (% of Sample)	Singletons (% of Sample)
<i>Algebra</i>		
3	48.26	47.51
4	39.14	39.77
5	12.60	12.72
<i>Geometry</i>		
3	47.58	47.63
4	39.72	39.24
5	12.70	13.13

Without going into further detail, we should note that performance was also compared within subgroups. No differences in mathematical achievement were observed between subgroups based on zygosity (monozygotic vs. dizygotic) or type (single-sex vs. opposite-sex). Thus, the study showed that zygosity does not affect scholastic achievement in algebra ( $\chi^2 = 3.97$ ,  $df = 4$ ,  $p = 0.41$ ) or geometry ( $\chi^2 = 6.78$ ,  $df = 4$ ,  $p = 0.14$ ). For more details, please see Zyrianova, 2009 a, b.

Table 2 provides data on the USE math scores of twin partners (4.8% of sample) and of all students (both twins and singletons, i.e., 100% of sample) who took the USE. The division of scores into subgroups, as well as the designation of the score levels (minimal, low, medium, and high) is based on the classification developed by the Federal Institute of Pedagogical Measurement, derived from USE math scores.

**Table 2.** Distribution of high school graduates by performance level (percent of sample) based on USE math scores

Performance Level	Twins (% of Sample)	All USE Takers (% of Sample)
Minimal	10.50	12.42
Low	65.58	65.80
Medium	22.79	20.84
High	1.13	0.94

Since no data are available on the USE math scores of only singleton students, it was not possible to conduct a statistical analysis of intergroup differences. However, by comparing the USE scores of twins vs. all USE takers, we can show that the mathematical achievement of twins is at least as strong as that of singletons. Thus, only 10.5% of twins received minimal USE scores (fewer than singletons with

minimal scores), while the percentage of twins who received medium or high USE scores was higher than that of singletons.

Both measures of academic performance by students (school grades and test scores) indicate that there are no systematic differences in the mathematical achievement of singletons and twins in grades 8–11.

### *Sex differences in math grades*

The distribution of boys vs. girls (percent of sample) in the sample of Study 1 was 46.73 vs. 53.28. Table 3 summarizes data on the algebra and geometry grades of twins and singletons, broken down by sex.

**Table 3.** Distribution of boys vs. girls (percent of sample) by final grade in algebra and geometry (grades 8–11)

Final Grade	Twins		Singletons	
	Boys (% of Sample)	Girls (% of Sample)	Boys (% of Sample)	Girls (% of Sample)
Algebra				
3	57.60	40.10	57.07	37.82
4	31.64	45.68	34.32	45.22
5	10.76	14.22	8.61	16.96
Geometry				
3	58.59	38.00	57.19	38.02
4	30.76	47.52	33.77	44.74
5	10.65	14.48	9.04	17.24

Girls earned higher math grades both in the twin group and in the singleton group. The differences were significant for both algebra and geometry:  $\chi^2 = 74.08$ ,  $p < 2.2e-16$  for boy twins vs. girl twins in algebra,  $\chi^2 = 89.02$ ,  $p < 2.2e-16$  for boy singletons vs. girl singletons in algebra,  $\chi^2 = 99.695$ ,  $p < 2.2e-16$  for boy twins vs. girl twins in geometry, and  $\chi^2 = 84.339$ ,  $p < 2.2e-16$  for boy singletons vs. girl singletons in geometry.

Students with 5s in both algebra and geometry included 304 girls (63.60%) and 181 boys (36.4%). Considering the fact that the number of boys and girls in the sample was almost equal (49.53% vs. 50.47%), we can conclude that the relative number of girls with top grades was higher. The size effect also attests to sex differences ( $d = -0.33$  for algebra and  $d = -0.41$  for geometry).

Since the average age of participants in Study 1 was 1.5 years younger than in Study 2, the performance of the graduating class (grade 11, mean age of 16.6) was analyzed separately. These results exhibited the same patterns as data for all high school students (no difference between twins vs. singletons in mathematical achievement and higher math grades for girls than for boys). The difference in the final grades of twins vs. singletons in 11th grade was insignificant in both algebra ( $\chi^2 = 0.26$ ,  $p = 0.877$ ) and geometry ( $\chi^2 = 0.23$ ,  $p = 0.893$ ). Girls had stronger scholastic performance than boys in both algebra ( $\chi^2 = 74.08$ ,  $p < 2.2e-16$ ) and geometry ( $\chi^2 = 99.695$ ,  $p = 2.2e-16$ ). Since the results for this subgroup did not differ from the results for the entire sample of Study 1, only data for all high school students are described henceforth.

**Sex differences in USE math scores**

A comparison of the mean USE test scores of boys vs. girls yields completely different results. Boys score much higher than girls on USE math sections (Table 4) and have higher overall test scores on average (47.56 vs. 46.82,  $p < 0.001$ ), but the effect size is very close to zero ( $d = 0.05$ ).

**Table 4.** Sex differences in USE math scores: mean standard deviation, Cohen's  $d$ 

Sex	Mean	Standard Deviation	$t$ -Criterion	$F$ -Ratio	$d$
Boys	47.56	15.46	3.59	1.10	0.05
Girls	46.82	14.76	$p < 0.001$	$p < 0.001$	

Sex differences in math ability can be observed not only in comparing average scores, but also in analyzing variability (boys exhibit significantly higher variation in mathematical achievement). The difference in variability is also significant when comparing extreme groups (previously described in detail, see Chertkova & Egorova, 2013 and Chertkova & Pyankova, 2014).

It was not possible to isolate extreme subgroups based on school grades in Study 1: The aforementioned tendency of teachers to avoid the lowest grades means that there are virtually no 2s in math, while almost half of the students get 3s.

**Table 5.** Number of boys and girls in extreme groups of USE math scores

USE Score	Number in group		% of Extreme Group		% of Entire Sample	
	Boys	Girls	Boys	Girls	Boys	Girls
<i>Criterion for group selection: <math>M \pm 2\sigma</math></i>						
Low USE scores	192	260	42.48%	57.52%	2.01%	2.04%
High USE scores	224	162	58.03%	41.97%	2.34%	1.27%
<i>Criterion for group selection: 0.5% tails of distribution</i>						
Low USE scores	47	54	46.53%	53.47%	0.49%	0.42%
High USE scores	48	31	60.76%	39.24%	0.50%	0.24%

The distribution of mathematical achievement indicators in Study 2 is close to the normal distribution curve, which makes it possible to analyze the ratio of boys to girls in extreme groups. The sample size allowed us to isolate extreme subgroups using two types of criteria. First, two groups of high school students were selected whose USE math scores were at least 2 standard deviations away from the mean ( $M \pm 2\sigma$ , soft criterion). The group of lowest-performing students included twins who received a score of no more than 16 points on a scale of 0–100 (452 students). The group of highest-performing students included twins who received a score of at least 79 points (386 students). Second, students who received the lowest 0.5% and highest 0.5% of scores among all USE test takers were separated out (hard criterion). The former received a score of no more than 5 points (101 students); the

latter received a score of at least 90 points (79 students). Table 5 summarizes the distribution of boys vs. girls in these extreme groups.

The ratio of boys to girls in the group of the lowest-performing students based on the soft criterion was proportional to the overall ratio of boys to girls in the entire sample (Study 2 had 42.84% boys vs. 57.16% girls). The highest-performing group had significantly more boys than girls (58.3% vs. 41.97%,  $\chi^2 = 37.06$ ,  $p = 8.958e-09$ ). With the use of the hard criterion, the proportion of boys in either tail of the distribution was relatively higher ( $\chi^2 = 10.995$ ,  $p = 0.004$ ).

The ratio of boys to girls among the highest-performing students based on the soft criterion was 1.38. The ratio of boys to girls among the highest-scoring students based on the hard criterion was 1.55.

For the entire sample, 2.34% of boys were in the highest-performing group based on the soft criterion and 0.50% based on the hard criterion. The percentage of girls in the top-performing group was 1.27% based on the soft criterion and 0.24% based on the hard criterion. Thus, there were twice as many boys as girls in the top tail of the USE math score distribution curve (ratio of 1.84 based on one criterion and 2.08 based on the other). The harder the selection criterion, the greater the advantage of boys over girls in the highest-scoring group.

### ***Sex differences in academic achievement of twins from opposite-sex pairs***

Analysis of the twin sample in the study made it possible to compare the performance of boys and girls from opposite-sex twin pairs and assess whether environmental factors related to having a co-twin of the opposite sex affect sex differences in mathematical ability.

The sample in Study 1 included 254 opposite-sex twin pairs; the sample in Study 2 included 2,562 opposite-sex pairs. Table 6 presents data on scholastic mathematical achievement of boys and girls from opposite-sex pairs.

**Table 6.** Sex differences in math grades and USE scores for twins from opposite-sex pairs: descriptive statistics and Cohen's  $d$

Mathematical achievement	Boys		Girls		$\chi^2$	t	$d$
	M	SD	M	SD			
Algebra grades	3.52	0.72	3.77	0.73	14.630		-0.34
Geometry grades	3.52	0.73	3.72	0.71	14.663		-0.28
USE scores	47.85	14.79	47.85	14.69		0.006	0.00

Data from Study 1 show that practically all opposite-sex twins had similar grades in algebra and in geometry—i.e., those who do well in algebra do well in geometry and vice versa.

Girls had better grades in both subjects than boys ( $\chi^2 = 14.630$ ,  $p = 0.0006$  for algebra;  $\chi^2 = 14.663$ ,  $p = 0.0006$  for geometry), which is in line with the results obtained for the entire sample. However, effect size for opposite-sex twin pairs was somewhat lower than for the entire sample ( $d = -0.28$  vs.  $d = -0.41$ ).

A comparison of the USE math scores of opposite-sex twins (Study 2) yields somewhat different results than those for the entire sample: Boys from opposite-sex

twin pairs had USE math scores comparable to boys from the entire sample (47.45 vs. 47.85,  $p = 0.26$ ); however, girls from opposite-sex pairs had significantly different scores (46.55 vs. 47.85,  $p < 0.001$ ): Girls from opposite-sex pairs performed much better on the USE math section than girls from the entire sample.

Effect size ( $d$ ), which showed a slight advantage for boys in the entire sample, indicates that there were no sex differences in mathematical achievement in opposite-sex twin pairs ( $d = 0.00$ ).

### ***Positive mathematical self-concept***

To graduate, high school students take two mandatory exams (mathematics and Russian language) and several optional exams in subjects they select. The choice of additional exams is not completely free and depends on the major a student wants to pursue in college. Educational institutions that offer STEM programs require applicants to take exams not only in math but also in physics. For this reason, USE test takers who select an optional physics exam in addition to the mandatory sections most likely intend to study mathematics in college and then enter STEM fields. Based on this reasoning, an optional physics exam can be viewed as an indicator of positive self-concept and self-confidence in mathematics.

A subgroup of twins who took the USE physics exam was selected from the entire sample of Study 2. Table 7 lists the mean scores that these students received on the USE math section.

**Table 7.** USE math scores of students who took an optional physics exam: descriptive statistics and Cohen's  $d$

	Mean	SD	$t$ -Criterion	$F$ -Ratio	$d$
Boys	52.87	14.99	-4.54	1.05	-0.13
Girls	54.83	14.66	$p < 0.001$		

Among students who expressed interest in STEM fields, there were significantly more boys than girls. At the same time, the girls in this subgroup scored higher on the USE math section than the boys (average score of 54.83 for girls and 52.87 for boys,  $p < 0.001$ ). Effect size ( $d$ ) was -0.13, which suggests a small sex difference; at the same time, for the entire sample, effect size was smaller and opposite in direction. In other words, data for the entire sample showed a certain advantage for boys, but girls outperformed boys in mathematical achievement in the subgroup of students with a higher positive mathematical self-concept.

### **Discussion**

Data obtained through a comparison of the school grades and USE scores of girls and boys were in line with the results of most studies: Girls had higher algebra and geometry grades, but slightly lower USE scores.

Let us note above all that differences between the USE scores of boys and girls were small: There was a statistically significant mean difference, but negligible effect size. Data on changes in the mathematical achievement of boys and girls over the past decade show that sex differences have been decreasing. Thus, in 2003, there

was an 11-point gap in the math scores of Russian eighth-grade boys and girls on the Trends in International Mathematics and Science Study (TIMSS). The average difference in the scores of eighth-graders from 34 countries was 8.6 points; raw score gaps ranged from  $-27$  to  $+29$  (Nosic et al., 2009). In 2011, the gap in the scores of Russian eighth-graders was only 1 point and in favor of the girls (Mullis et al., 2012). Shrinking sex differences in mathematical achievement can be observed in various countries; the advantage of boys in a meta-analysis was small (Cohen's  $d = 0.05$ ) and coincided with data obtained in our research ( $d = 0.06$ ).

Our study demonstrated much larger differences in the school grades of boys and girls. Effect sizes in final grades in algebra and geometry equaled  $-0.33$  and  $-0.41$ , respectively. A meta-analysis of sex differences in school grades (Voyer & Voyer, 2014) indicated that girls in older grades performed better in math, but the difference between boys and girls was smaller: Cohen's  $d = -0.18$ .

There are no unequivocal explanations for the opposite directions in sex differences based on various measures of mathematical achievement, despite the hundreds of studies conducted. The dynamics are not as simple as they might appear at first glance. There is no empirical evidence for "obvious" explanations (e.g., that teachers favor diligent and hard-working girls and give them better grades, while boys have better math abilities on average and therefore outperform girls on tests where a teacher's personal feelings do not have an effect). For example, data from USA show that teachers encourage boys more during lessons, call on boys more frequently, and respond to boys' questions more often in conditions of comparable initiative by boys and girls (Jones & Dindia, 2004). The interpretation of sex differences requires consideration of more complex mechanisms related to self-concept and extrinsic/intrinsic motivation. These interrelated psychological traits—each of which is a complex construct—reveal many direct and indirect links with sex differences in both school grades and test scores. The interplay of beliefs, motivations, learning styles, and academic achievement is also vital to understanding sex differences (Lee et al., 2014; Muis, 2014; Gaspard et al., 2015; Guo et al., 2015).

Our research also supported the link with mathematical self-concept: Girls who selected an optional USE physics exam (i.e., have a positive math self-concept and intend to pursue degrees linked with math) received better USE scores than boys overall, as well as better than boys who plan to continue studying mathematics. It appears that a positive math self-concept requires a higher degree of security from girls than from boys. The only study we have seen that takes an analogous approach yielded similar results (Korpershoek et al., 2011). As in our study, high school students in the Netherlands selected sections to take as part of their final school examination. Less than 12% of students took math, physics, and chemistry exams: girls were in the minority in this group, but they outperformed boys on the math exam. In other words, among boys and girls with the same math grades, boys appear to have higher confidence in their readiness to enter STEM fields and are more likely to pursue STEM degrees than girls.

Another comparison of boys and girls in our study was conducted using the group of opposite-sex DZ twins. Many studies of twins and singletons indicate that twins lag in cognitive development: they have lower intelligence scores and weaker academic performance (for example, Deary et al., 2005; Ronalds, et al., 2005; Christensen et al., 2006; Voracek & Haubner, 2008; Behrman, 2015). However, with age these dif-



ferences decrease significantly or even disappear (Deary et al., 2006; Webbink et al., 2008; Calvin et al., 2009; Eriksen et al., 2012). Our study showed that high-school-age twins do not have inferior school grades or USE test scores than singletons.

Twin samples allow quasi-experimental designs that are impossible with singleton samples. Thus, a comparison of boys and girls from opposite-sex twin pairs equalizes a range of parameters including age and certain family and school environment indicators (socioeconomic status, personality traits of parents, parenting styles, etc.) This significantly cuts down on characteristics that can affect the development or absence of sex differences.

The girls from opposite-sex DZ pairs in our study showed no difference from their twin brothers in mathematical achievement based on USE scores. This result suggests that the concept of gender-differentiated parental expectations (expectations of higher mathematical achievement by boys), which has been widely discussed over the last decade and a half, is unlikely to be a significant moderator of mathematical achievement. Two hypotheses with contradictory implications can be put forth regarding the results obtained in the study (the parity in mathematical achievement of boys and girls from opposite-sex twin pairs).

The first relates to the similar environment of DZ pairs: Twins spend a lot of time together and share each other's interests, which leads to comparable mathematical achievement. This hypothesis requires further study at the very least, since DZ twins, particularly opposite-sex twins, tend towards divergence rather than convergence of activities and interests. Furthermore, this hypothesis does not appear convincing in light of data on differences in the mathematical abilities of siblings regardless of birth order (Cheng et al., 2012), as well as data on low mathematical achievement of adopted children (van Ijzendoorn et al., 2005).

The second hypothesis hinges on the link between mathematical ability and prenatal testosterone levels (twin testosterone transfer hypothesis). There is a theory that during the prenatal period, having an opposite-sex co-twin can change the level of prenatal testosterone, resulting in differentiated brain structure and masculinization of girls (Tapp et al., 2011; Ahrenfeldt et al., 2015). This means that girls from opposite-sex pairs are more likely to pursue activities linked to the development of spatial abilities (Berenbaum et al., 2012; Constantinescu & Hines, 2012) and have fewer differences from boys in mathematical ability and mathematical achievement (as found in our study). Thus, biological factors related to the formation of mathematical ability could be linked with sex differences as well. Moreover, they could contribute significantly to sex difference indicators such as dispersion of mathematical ability (which is higher for boys in our study as well as in other studies) and the greater proportion of boys in the highest-achieving groups (our study showed that the ratio of boys to girls in the highest-achieving groups was approximately 2:1).

## **Conclusion**

Academic achievement in math differs for boys and girls, but the direction of difference varies depending on how achievement is measured. Girls have higher school grades while boys have higher USE test scores.

The number of boys in the right tail of distribution is greater than the number of girls.

Within the group of high school students with a positive mathematical self-concept, girls outperform boys in mathematical achievement.

Girls from opposite-sex DZ pairs show better mathematical achievement than singleton girls in their age group and do not differ in mathematical ability from boys.

## References

- Ahrenfeldt, L., Inge, P., Wendy, J., & Christensen, K. (2015). Academic performance of opposite-sex and same-sex twins in adolescence: A Danish national cohort study. *Hormones and Behavior*, *69*, 123–131. doi: 10.1016/j.yhbeh.2015.01.007
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty and student learning. *Education Policy Analysis Archives*, *10*(18), 1–74. Retrieved from <http://epaa.asu.edu/epaa/v10n18>
- Behrman, J. R. (2015). Twin studies in demography. In J. D. Wright (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition) (pp. 703–709). Oxford, England: Elsevier. doi: 10.1016/B978-0-08-097086-8.31130-8
- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, *11*, 169–232. doi: 10.1017/S0140525X00049244
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210*, 1262–1264. doi: 10.1126/science.7434028
- Benbow, C. P., & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal*, *19*, 598–622. doi: 10.3102/00028312019004598
- Berenbaum, S. A., Bryk, K. L., & Beltz, A. M. (2012). Early androgen effects on spatial and mechanical abilities: evidence from congenital adrenal hyperplasia. *Behavioral Neuroscience*, *126*, 86–96. doi: 10.1037/a0026652
- Byrne, B. M. (1996). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Eds.), *Handbook of Self-Concept* (pp. 287–316). New York: Wiley.
- Chamorro-Premuzic, T., Harlaar, N., Greven, C. U., & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence*, *38* (4), 385–392. doi: 10.1016/j.intell.2010.05.002
- Calvin, C., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2009). Is there still a cognitive cost of being a twin in the UK? *Intelligence*, *37*(3), 243–248. doi: 10.1016/j.intell.2008.12.005
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: a changing landscape. *Psychological Science in the Public Interest*, *15*, 75–141. doi: 10.1177/1529100614541236
- Cheng, C-C. J., Wang, W-L., Sung, Y-T., Wang, Y-C., Su, S-Y., & Li, C-Y. (2013). Effect modification by parental education on the associations of birth order and gender with learning achievement in adolescents. *Child: Care, Health and Development*, *39*(6), 894–902.
- Chertkova, Y. D., & Egorova, M.S. (2013). Polovye razlichija v matematicheskikh sposobnostjah [Sex differences in mathematical abilities]. *Psikhologicheskie Issledovaniya [Psychological Studies]*, *6*(31), 12. Retrieved from <http://psystudy.ru>
- Chertkova, Y. D., & Pyankova, S. D. (2014). Akademicheskaja uspevaemost bliznecov i odinochnorozhdennyh detej: Kross-kulturnoe issledovanie [Sex differences in academic achievement depending on the professional self-determination of schoolchildren]. *Psikhologicheskie Issledovaniya [Psychological Studies]*, *7*(38), 10. Retrieved from <http://psystudy.ru>
- Christensen, K., Peterson, I., Skytthe, A., Herskind, A. M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: Follow-up study. *British Medical Journal*, *333*, 1095–1097. doi: 10.1136/bmj.38959.650903.7C

- Constantinescu, M., & Hines, M. (2012). Relating prenatal testosterone exposure to postnatal behavior in typically developing children: Methods and findings. *Child Development Perspectives*, 6, 407–413. doi: 10.1111/j.1750-8606.2012.00257.x
- Cucina, J. M., Peyton, S. T., Su, C., & Byle, K. A. (2016). Role of mental abilities and mental tests in explaining high-school grades. *Intelligence*, 54, 90–104. doi: 10.1016/j.intell.2015.11.007
- Deary, I. J. (2006). Educational performance in twins is no different from that seen in singletons by adolescence. *British Medical Journal*, 333, 1080–1081. doi: 10.1136/bmj.39037.543148.80
- Deary, I. J., Pattie, A., Wilson, V., & Whalley, L. J. (2005). The cognitive cost of being a twin: Two whole-population surveys. *Twin Research and Human Genetics*, 8, 376–383. doi: 10.1375/twin.8.4.376
- Ding C.S, Song K., & Richardson L.I. (2006). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies*, 40, 3, 279–295. doi: 10.1080/00131940701301952
- 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. doi: 10.1037/a0018053
- Eriksen, W, Sundet, J. M, & Tambs, K. (2012). Twin-singleton differences in intelligence: A register-based birth cohort study of Norwegian males. *Twin Research and Human Genetics*, 15(5), 649–655. doi: 10.1017/thg.2012.40
- Gaspard H., Dicke A-L, Flunger B., Schreier B., Häfner I., Trautwein U., & Nagengast B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, 107, 3, 663–677. doi: 10.1037/edu0000003
- Guo J., Marsh H.W., Parker P.D., Morin A.J.S., & Yeung A.S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, 37, 161–168. doi: 10.1016/j.lindif.2015.01.008
- Hofer, M., Kuhnle, C., Kilian, B., & Fries, S. (2012). Cognitive ability and personality variables as predictors of school grades and test scores in adolescents. *Learning and Instruction*, 22(5), 368–375. doi: 10.1016/j.learninstruc.2012.02.003
- Hyde J. S., Fennema E., & Lamon S. J. (1990). Gender differences in mathematics Performance: A meta-analysis. *Psychological Bulletin*, 107(2), 139–155. doi: 10.1037/0033-2909.107.2.139
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321, 494–495. doi: 10.1126/science.1160364
- Hyde, J. S., Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, 314, 599–600. doi: 10.1126/science.1132154
- Jones S.M., & Dindia K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, 74, 4, 443–471. doi: 10.3102/00346543074004443
- Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin*, 105, 198–214. doi: 10.1037/0033-2909.105.2.198
- Kling, K., Hyde, J., Showers, C., & Buswell, B. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, 125, 470–500. doi: 10.1037/0033-2909.125.4.470
- Korpershoek, H., Kuyper, H., van der Werf, G., & Bosker R. (2011). Who succeeds in advanced mathematics and science courses? *British Educational Research Journal*, 37(3), 357–380. doi: 10.1080/01411921003671755
- Lee, W., Lee, M-J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology*, 39, 86–99. doi: 10.1016/j.cedpsych.2014.02.002
- 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. doi: 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. *British Journal of Educational Psychology*, 79, 389–407. doi: 10.1348/000709908X354609
- Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, 1, 316–345. doi: 10.1111/j.1745-6916.2006.00019.x
- Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science*, 17, 194–199. doi: 10.1111/j.1467-9280.2006.01685.x
- Luo, Y.L.L., Kovas, Y., Haworth, C.M.A., & Plomin, R. (2011). The etiology of mathematical self-evaluation and mathematics achievement: Understanding the relationship using a cross-lagged twin study from ages 9 to 12. *Learning and Individual Differences*, 21, 6, 710–718. doi: 10.1016/j.lindif.2011.09.001
- McClure, J., Meyer, L.H., Garisch, J., Fischer, R., Weir, R.F., & Walkey F.H. (2011). Students' attributions for their best and worst marks: Do they relate to achievement? *Contemporary Educational Psychology*, 36, 71–81. doi: 10.1016/j.cedpsych.2010.11.001
- Marsh, H. W. (1986). Verbal and math self-concepts: an internal/external frame of reference model. *American Educational Research Journal*, 23, 129–149. doi: 10.3102/00028312023001129
- Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2, 77–172. doi: 10.1007/BF01322177
- Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 154–157. doi: 10.1207/s15326985ep3403\_2
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1, 133–163. doi: 10.1111/j.1745-6916.2006.00010.x
- Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of East and West German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology*, 29(3), 264–282. doi: 10.1016/S0361-476X(03)00034-1
- Marsh, H.W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., & Abdelfattah, F., Jansen, M. (2015). Dimensional Comparison Theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, 35, 16–32. doi: 10.1016/j.learninstruc.2014.08.005
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effect models of causal ordering. *Child Development*, 76, 397–416. doi: 10.1111/j.1467-8624.2005.00853.x
- Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal*, 34, 691–720. doi: 10.3102/00028312034004691
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129–1167. doi: 10.3102/0034654309337522
- Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal I/E model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315–1346. doi: 10.3102/0002831211419649
- Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, 84, 591–611. doi: 10.1111/bjep.12047

- Morosanova, V. I., Fomina, T. G., Kovas, Yu. V. (2014). Vzaimosvjaz reguljatornyh, intellektualnyh i kognitivnyh osobennostej uchashhihsja s matematicheskoj uspešnostju [The relationship between regulatory, intellectual and cognitive characteristics in students who are successful in mathematics]. *Psichologičeskie Issledovaniya [Psychological Studies]*, 7(34), 11. Retrieved from <http://psystudy.ru>
- Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research Fall*, 74(3), 317–377. doi: 10.3102/00346543074003317
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement.
- Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, 106(4), 1170. doi: 10.1037/a0036307
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, ... Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academies of Science*, 106, 10593–10597. doi: 10.1073/pnas.0809921106
- Novikova, M. A., & Kornilova, T. V. (2012). Samoocenka intelekta v strukturnyh svjazjah s psihometričeskim intelektom, ličnostnymi svojstvami i akademicheskoj uspevaemostju [Intelligence self-evaluation in structural links with psychometric intelligence, personality traits, academic achievements and gender]. *Psichologičeskie Issledovaniya [Psychological Studies]*, 5(23), 2. Retrieved from <http://psystudy.ru>
- 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(3), 645–662. doi: 10.1037/edu0000012
- Ronalds, G. A., De Stavola, B. L., & Leon, D. A. (2005). The cognitive cost of being a twin: Evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. *British Medical Journal*, 331, 1306. doi: 10.1136/bmj.38633.594387.3a
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137. doi: 10.1016/j.intell.2015.09.002
- Seaton, M., Parker, P., Marsh, H.W., Craven, R. G., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: Juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 34(1), 49–72. doi: 10.1080/01443410.2013.825232
- Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., & Conley, A. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high- and low-stakes test achievement. *Learning and Individual Differences*, 39, 49–63. doi: 10.1016/j.lindif.2015.03.002
- Skaalvik, E., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist*, 37, 233–244. doi: 10.1207/S15326985EP3704\_3
- Spinath, F. M., Spinath, B., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*, 22(3), 211–229. doi: 10.1002/per.677



- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology*, 76, 463–480. doi: 10.1348/000709905X50906
- Syzmanowicz, A., Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta-analyses. *Learning and Individual Differences*, 21, 493–504. doi: 10.1016/j.lindif.2011.07.001
- Tapp, A. L., Maybery, M. T., & Whitehouse, A. J. O. (2011). Evaluating the twin testosterone transfer hypothesis: A review of the empirical evidence. *Hormones Behavior*, 60, 713–722. doi: 10.1016/j.yhbeh.2011.08.011
- Trautwein, U., Ludtke, O., Marsh, H. W., Koller, O., & Baumert, J. (2006). Tracking, grading and student motivation: Using group composition and status to predict self-concept and interest in ninth grade mathematics. *Journal of Educational Psychology*, 98, 788–806. doi: 10.1037/0022-0663.98.4.788
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relations between self-beliefs and academic achievement: A systematic review. *Educational Psychologist*, 39, 111–133. doi: 10.1207/s15326985ep3902\_3
- van Ijzendoorn, M.H., Juffer, F., & Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin*, 131(2), 301–316. doi: 10.1037/0033-2909.131.2.301
- Vedel, A. (2016). Big Five personality group differences across academic majors: A systematic review. *Personality and Individual Differences*, 92, 1–10. doi: 10.1016/j.paid.2015.12.011
- Voracek, M., & Haubner, T. (2008). Twin-singleton differences in intelligence: A meta-analysis. *Psychological Reports*, 102, 951–962. doi: 10.2466/pr0.102.3.951-962
- Voyer, D., & Voyer, S.D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. doi: 10.1037/a0036620
- Wai J., Cacchio M., Putalla, M., & Makel, M.C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence*, 38(4), 412–423. doi: 10.1016/j.intell.2010.04.006
- Wang, M-T., & Degol, J. L. (2016). Gender gap in Science, Technology, Engineering, and Mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 1–22. doi: 10.1007/s10648-015-9355-x
- Webbink, D., Posthuma, D., Boomsma, D. I., de Geus, E. J.C., & Visscher, P. M. (2007). Do twins have lower cognitive ability than singletons? *Intelligence*, 36(6), 539–547. doi: 10.1016/j.intell.2007.12.002
- Xu, M. K., Marsh, H.T. W., Hau, K-T, Ho, I. T., Morin, A. J. S., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. *Journal of Educational Psychology*, 105(2), 489–503. doi: 10.1037/a0031333
- Zyrianova, N. M. (2009). Akademicheskaja uspešnost bliznecov i ih odinochnorozhdennyh sverstnikov. Chast 1 [Academic achievement of twins and their single-born peers. Part 1]. *Psikhologičeskie Issledovaniya [Psychological Studies]*, 4(6). Retrieved from <http://psystudy.ru>
- Zyrianova, N. M. (2009). Akademicheskaja uspešnost bliznecov i ih odinochnorozhdennyh sverstnikov. Chast 2 [Academic achievement of twins and their single-born peers. Part 2]. *Psikhologičeskie Issledovaniya [Psychological Studies]*, 5(7). Retrieved from <http://psystudy.ru>

Original manuscript received November 28, 2015

Revised manuscript accepted April 02, 2016

First published online September 30, 2016