High mathematics and reading performance

Schwabe, Inga; Meelissen, M.R.M.; Punter, R.A.; van den Berg, S.M.

Published in:
Talent Development & Excellence

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

High mathematics and reading performance: How important are environmental influences?

I. Schwabe[1,2,*], M. R. M. Meelissen[1], R. A. Punter[1], and S. M. van den Berg[1]

Abstract: Earlier findings of international comparisons on school achievement are often interpreted to mean that there is only a small percentage of excellent students in the Netherlands. Inspired by research in behaviour genetics, it was investigated whether Dutch high-scoring children are less sensitive to environmental influences than the non-high-scoring students. To test this, the reading and mathematics scores from high-scoring and non-high-scoring students participating in the *Programme for International Student Achievement* (PISA) 2012, the *Trends in International Mathematics and Science Study* (TIMSS) 2011 and the *Progress in International Reading and Literacy Study* (PIRLS) 2011 were analyzed. Contrary to our expectations, the results suggest that high-scoring children are as sensitive to school influences as are non-high-scoring students, but more sensitive to the influence of individual socioeconomic status.

Keywords:
mathematics performance, reading performance, PISA, TIMSS, high-scoring students

Introduction

When the performance of Dutch students on international comparisons such as the *Programme for International Students Achievement* (PISA) and the *Trends in International Science Study* (TIMSS) is compared to the performance of students from other participating countries, a striking result is that the variance of test scores is very small: While Netherlands' weakest students perform better than the weakest students from all other countries, the highest-scoring students are outperformed by the top students from Asian and other Western countries (see, e.g., Meelissen et al., 2012; van der Steeg, Vermeer, & Lanser, 2011). These findings are often presented as underperformance of the most talented students (e.g., van der Steeg, Vermeer, & Lanser, 2011) and interpreted to indicate that Dutch mathematical education is better tailored to the weaker than the mathematically talented students.

From a behaviour genetics standpoint, a situation in which differences in test scores could be explained mainly by genetic differences rather than random situational factors such as being at the mercy of a particular teacher would imply that the Dutch education is ideal for every child – regardless of environmental conditions (see Shakeshaft et al., 2013 for a similar argument). This line of reasoning would also imply that, if indeed, in primary education, mathematically talented children are not nurtured to their full performance, differences in test scores should be more affected by environmental factors than the performance of average or weak students of the same age (i.e., genotype by environment interaction). Using the item answers of 2010 12-year-old Dutch twin pairs on the mathematical subscale of a national achievement test administered in the final year of primary education, Schwabe, Boomsma and van den Berg (2017) investigated such an interaction between mathematical talent (defined as a child's genotypic value) and

[1] *Department of Research Methodology, Measurement and Data Analysis (OMD), University of Twente, the Netherlands*
[2] *Department of Methodology and Statistics, Tilburg University, the Netherlands*
** Corresponding author: PO Box 90153, 5000 LE Tilburg, the Netherlands*
Email: I.Schwabe@uvt.nl

ISSN 1869-0459 (print)/ ISSN 1869-2885 (online)

http://www.iratde.org

environmental influences. As hypothesized, results suggested that environmental influences were relatively more important in explaining individual differences in students with mathematical talent than in students without special talent for high mathematics performance.

As, arguably, the performance of students with mathematical talent is more heavily influenced by random environmental influences such as being at the mercy of a particular teacher than the performance of students without such a talent, these findings might be interpreted to suggest that the Dutch educational system, indeed, is inappropriate for mathematically talented students. Earlier research provides support for this interpretation. For example, research by Mooij, Hoogeven, Driessen, Hell and Verhoeven (2007) suggests that Dutch talented students are not challenged enough and do not get enough possibilities to develop their talent. Other researchers found that talented children need a different stimulation than average or weak performing children of the same age (e.g., Cigman, 2006; van Houten, 2009). For example, van Houten (2009) argues that the Netherlands' current educational system is not suitable for talented children, because they may have difficulty with the slow pace of the class and get bored. As a result, they lose their motivation and perform less well than their capacities allow them to.

However, before drawing any conclusions, it is crucial to investigate whether the results of the twin sample used by Schwabe, Boomsma and van den Berg (2017) can be generalized to the general population of Dutch students. In their research, they applied the twin design, a commonly used method in the field of behaviour genetics. Using data from identical (monozygotic, MZ) as well as non-identical (dizygotic, DZ) twins, the twin design makes it possible to estimate how much of the variance in an observed trait (e.g., mathematical ability) is explained by genetic influences and common-environmental influences that are shared by the same family. Variance that is not explained by these influences is attributed to unique-environmental influences that are not shared by the same twin pair. However, the twin sample that was used in Schwabe et al. (2017) is not necessarily representative of the general student population – for example, singletons differ significantly from twin pairs with regard to birth conditions (Martin et al., 2010). Furthermore, the data that were used originate in the Netherlands Twin Register (NTR, Boomsma et al., 2002) which is a voluntary register and might therefore introduce bias due to self-selection.

Furthermore, it is interesting to investigate whether the results obtained by Schwabe et al. (2017) can be generalized to a different domain than mathematics. Although, generally, Dutch highest-scoring students performed better on the reading than on the mathematics domain in international comparisons, they were outperformed by the top students from other participating countries on the reading domain as well. Twin research in the United Kingdom has shown that approximately one half of the observed correlation in mathematics and reading scores is due to shared genetic effects (i.e., the same genes are important for both, reading and mathematics performance) (Donnelly, Plomin, & Spencer, 2014). Given this association, we hypothesize that, for reading scores too, environmental influences are more important for high-scoring students than for low- or average-scoring students.

In this article, we used the mathematics and reading performance of Dutch students in the TIMSS, the PISA and the Programme for International Reading Literacy Study to investigate whether the results of Schwabe et al. (2017) can be replicated to 1) a different domain (i.e., reading performance) and 2) the general student population. In order to investigate whether high-scoring children are more sensitive to environmental influences, students were divided in two groups: high-scoring students and low-scoring students. High-scoring students were defined as the 20% best scoring children in the respective sample (Renzulli, Reis, & Smith, 1981) and low-scoring students were defined as the 20% lowest scoring students in the respective sample. We then investigated whether the relationship between environmental influences and test scores is different for the high-scoring than for the low-scoring group of students. In the research by Schwabe et al. (2017), genetic and environmental influences were parametrized as latent (i.e., unmeasured) variables. Therefore, no conclusion could be drawn on the nature and importance of *specific* environmental influences that are more important for high-scoring students. Here, two types of environmental influences were investigated: the individual socio-economic status (SES) and the effect of the school. Thus, specifically, it was investigated whether the relationship between school influences and test scores (research question I) and the relationship between SES and test scores (research question II) is mediated by the grouping of students. For a better understanding, these two research questions are graphically displayed in Figure 1. These influences were chosen, as research shows that these belong to the most important factors in explaining individual differences in test scores (see, e.g., Baharudin & Luster, 1998; Eamon, 2005; Hochschild, 2003). We furthermore chose to use relatively broad environmental influences in order to study whether, indeed, there is a difference between the high-scoring and low-scoring children in the first place. In a follow-up study, we could then later determine which specific environmental influences are more or less important for the high-scoring students.

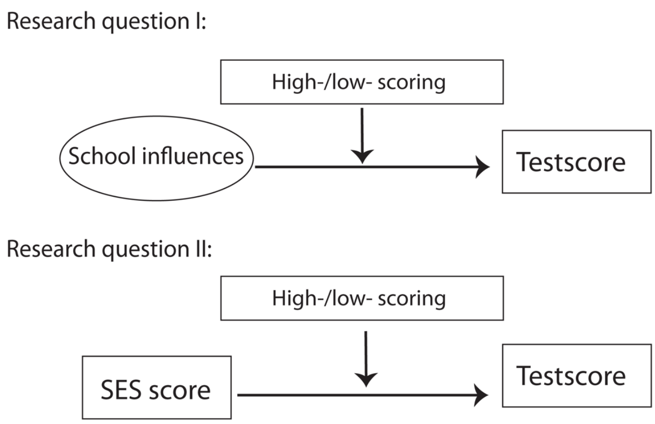

Figure 1. Graphic display of the research questions

Materials and Methods

In this study, we examined the test scores of Dutch pupils participating in PISA 2012, TIMSS 2011 and PIRLS 2011. The aim of PISA is to assess the skills of 15-year-olds in the domains of mathematics, reading and science. In TIMSS, the knowledge of 10-year-olds is tested in the domains of mathematics and science and PIRLS assesses the reading ability of 10 year-olds. This research investigated the mathematics and readings scores of 15-year-olds (PISA) and ten-year-olds (TIMSS and PIRLS). Table 1 provides a summary of all samples.

Table 1.
Summary of all samples

Sample	Domain	Age of the students	*N* schools	*N* students
PISA 2012	Mathematics and Reading	15-year-olds	179	4376
TIMSS 2011	Mathematics	10-year-olds	128	2584
PIRLS 2011	Reading	10-year-olds	128	3982

N = total.

Grouping of students

We defined high-scoring students as the 20% best scoring students and low-scoring students as the 20% lowest scoring students in the respective sample (PISA, TIMSS and PIRLS). A student's individual test score was used to determine whether a student belonged to the group of 'high-scoring' or 'low-scoring' students. In the PISA, TIMSS and PIRLS test, instead of a single test score, five so-called *plausible values* are used to determine a student's performance level. This is because these tests work with a rotated test design, in which each student is assigned to a test booklet. Individual test scores are therefore, initially, only comparable between students who were assigned to the same booklet. To make test scores assessed by different booklets comparable, a distribution of proficiency scores is constructed. In addition to properties of the test booklet, this distribution takes into account background characteristics of students. The plausible values are then five randomly drawn values from this distribution (Meelissen & Luyten, 2011; Von Davier, Gonzalez, & Mislevy, 2009). The division of students into the two groups of 'high-scoring' and 'low-scoring' was based on a student's individual test scores on all five plausible values that had to be above (high-scoring children) or below (low-scoring children) a critical value.

Analysis

For the analysis, we only used the data of the students that were divided into one of the two groups (i.e., high-scoring and low-scoring). An overview of these data per group can be seen in Table 2.

Table 2
Summary of the data, separately for each student group (i.e., low scoring and high scoring)

Sample	N schools		Average (SD)	
	H	L	H	L
PISA: mathematics	98	98	640 (38)	389 (41)
PISA: reading	104	99	622 (38)	373 (54)
TIMSS	116	118	604 (26)	470 (32)
PIRLS	128	130	611 (28)	474 (31)

N = total number of. SD = standard deviation. H = 20% highest-scoring students, L = 20% lowest-scoring students.

In order to investigate the first research question, we used multilevel analyses to compare the variance explained by school influences in the different student groups (i.e., high-scoring and low-scoring students). As, in each school, there were high-scoring as well as low-scoring children, it was decided to estimate variance explained by school influences (between-schools variance) separately for the group of high-scoring children and the group of low-scoring children. For every sample (PISA, TIMSS and PIRLS), following multilevel model was used:

$$y_{ij} = \beta_0 + b_{0i} + \epsilon_{ij}$$

where y_{ij} refers to the individual test score of student j from school i, β_0 denotes the population intercept and b_{0i} the school specific intercept. ϵ_{ij} refers to the error term, which was assumed normal and random. The model divides the total variance of all test scores into two independent components: between-schools variance and residual variance. In this research, we were specifically interested in between-schools variance. In order to determine whether between-schools variance was different for the two groups, a Wald-test was conducted for every sample (PISA: mathematics, PISA: reading, TIMMS and PIRLS).

As our second research question concerned an individual measure (SES), it was possible to answer this research question with one single multilevel analysis. Based on the grouping of students (i.e., high-scoring and non-high-scoring students), a dichotomous dummy variable was created (i.e., 1 = high scoring, 0 = low-scoring). Then, besides the random school effect and the main effects, an interaction (moderator) effect was estimated between individual SES score and the dummy variable:

$$y_{ij} = \beta_0 + \beta_1 * dummy_{ij} + \beta_2 * SES_{ij} + \beta_3 * (dummy_{ij} * SES_{ij}) + b_{0i} + \epsilon_{ij}$$

where y_{ij} refers to the individual test score of student j op school i, β_0 represents the population intercept and β_1, β_2 en β_3 denote regression coefficients for respectively the main effects and the interaction term. b_{0i} represent the school specific intercepts and ϵ_{ij} is the error term. To answer the research question, we are interested in the interaction effect, β_3. Depending on its direction, a significant β_3 would mean that the realtionship between test score and SES is stronger (positive β_3) or weaker (negative β_3) for the high-scoring children than for the low-scoring students.

In the PIRLS sample, SES represents the minimal education of a student's parents, measured on the following ordinal scale: 1) Some primary, lower secondary, or no school 2) Lower secondary, 3) Upper-secondary, 4) Post secondary but not university and 5) university or higher. To ease interpretation, this variable was summarized into three categories: 1) Lower secondary 2) Upper or post-secondary and 3) Bachelor HBO. These categories were interpreted as 'low SES', 'average SES' and 'high SES'. The education of the parents was measured by means of a questionnaire addressed to the parents. As not all parents completed this questionnaire, some of the SES data was missing. The education of the parents was known for a total of 2241 students (total PIRLS sample). For the 20% highest scoring students, the variable was known for a total of 577 students, of which the parents of 34 students were in the category 'lower secondary', 200 in the category 'upper or post-secondary' and 343 in the category 'Bachelor HBO'. For the 20% highest scoring students, the education of their parents was known for a total of 374 students, of which the parents of 85 belonged to the category 'lower secondary', 215 in the category 'upper or post-secondary' and 74 in the category 'bachelor HBO'.

The test scores were standardized in all analyses. Furthermore, in the analysis of the PISA sample, the SES variable was standardized. In each case, the standardization was done on population level, meaning that *all* available test scores (high-scoring students, low-scoring students, but average-scoring students) were used for the standardization. Furthermore, all analyses were conducted with weighted data, which makes results representative for Dutch students in the sixth year of Dutch primary school (i.e., between eight and eleven years; TIMSS and PIRLS) and fifteen-year-old Dutch students in Dutch secondary school (PISA). For the weighting of the PIRLS and TIMSS data, the publicly available SPSS (IBM Corp, 2011) code from the international coordination of the PIRLS and TIMSS sample was used. The PISA data were weighted by using the officially advised procedure by OECD (OECD, 2009).

Each analysis was conducted for all five plausible values. The reported variances and regression coefficients in the results section represent the averaged estimates over all five analyses. Due to the use of five different plausible values, the resulting standard errors had to be corrected (see, e.g., von Davier et al., 2009), which was done in this research by using the procedure that is officially advised by the OECD (OECD, 2003).

Results

Relationship between-school influences and test scores

The results of the first multilevel analyses can be found in Table 3 for the mathematics domain (PISA and TIMMS samples) and in Table 4 for the reading scores (PISA and PIRLS samples).

Table 3
Mathematics scores: Effect of school as function of score-group (PISA, TIMSS)

Selection based on testscore	Between-school variance (*SE*)	Between-school variance (*SE*)
Sample	*PISA*	*TIMSS*
20% highest-scoring students	0.01 (0.08)	0.00 (0.04)
20% lowest-scoring students	0.02 (0.11)	0.02 (0.11)

SE = Standard error.

Table 4
Reading scores: Effect of school as function of score-group (PISA, PIRLS)

Selection based on testscore	Between-school variance (*SE*)	Between-school variance (*SE*)
	PISA	PIRLS
20% highest-scoring students	0.01 (0.10)	0.00 (0.03)
20% lowest-scoring students	0.05 (0.21)	0.01 (0.08)

SE = Standard error.

Mathematics

It can be seen that the between-school variance was smaller for the group of high-scoring students than for the group of low-scoring students. Wald tests, however, showed that this difference was not significant for the PISA sample ($z = 0.15$, $p > 0.05$) nor the TIMSS sample ($z = 0.17$, $p > 0.05$).

Reading

Concerning reading performance, between-school variance was smaller for the high-scoring students than for the low-scoring students. Wald tests suggested that this difference was non-significant for the PISA sample ($z = 0.17$, $p > 0.05$) and the PIRLS sample ($z = 0.11$, $p > 0.05$).

Relationship between SES and testscores

Estimates for the so-called *fixed effects* (population intercept and regression coefficients for the main effects and the interaction effect) can be found in Table 5 for the mathematics scores (PISA sample) and in Table 6 for the reading scores (PISA and PIRLS samples). In the third column of this table, the estimated value of the parameters including standard error is displayed, in the fourth column the 95% confidence interval and in the last column the T- and p-values.

In order to interpret the above displayed parameter values more easily, results are graphically displayed in Figure 2 (PISA samples) and Figure 3 (PIRLS sample). In Figure 2, the SES and test scores are displayed for both groups of students (the gray dots belong the group of the 20% highest-scoring students) and the regression lines for the effect of SES on the test scores is shown for both groups separately. Figure 3 shows the boxplots for the different SES groups in the PIRLS sample. The results are displayed separately for both student groups.

Table 5
Mathematics scores: Results of the multilevel analyses for the PISA sample.

Sample	Parameter	Estimated value (*SE*)	95% Confidence-interval	*T*-value	*p*-value
PISA: Math-ematics	Intercept	-0.90 (0.02)	[-0.93;-0.87]	-56.97	<0.01
	Between-school variance	0.01 (0.10)	-	-	-**
	Dummy	1.81 (0.01)	[1.77;1.86]	81.96	<0.01
	SES	0.01 (0.01)	[-0.01;0.03]	1.23	0.22
	Dummy*SES	0.03 (0.02)	[0.01;0.06]	2.13	0.03

Note. *SE* = Standard error. ***The 95% confidence interval, T-value en p-value are not reported, as the inclusion of a random effect makes it difficult to calculate exact statistics (e.g., there is no agreed consensus yet on how to calculate the number of degrees and number of parameters, see e.g. Baayen, Davidson, & Bates, 2008).*

Table 6
Reading scores: Results of the multilevel analyses for the PIRLS sample.

Sample	Parameter	Estimated value (*SE*)	95% confidence-interval	*t*-value	*p*-value
PISA: reading	Intercept	-0.83 (0.02)	[-0.88;-0.79]	-39.80	<0.01
	Between-school variance	0.03 (0.17)	-	-	-**
	Dummy	1.70 (0.03)	[-1.64;1.76]	60.55	<0.01
	SES	0.01 (0.02)	[-0.01;0.03]	0.53	0.59
	Dummy*SES	0.04 (0.02)	[0.01;0.08]	2.73	<0.01
PIRLS	Intercept	-0.93 (0.03)	[-0.98;-0.87]	-33.48	<0.01
	Between-school variance	0.00 (0.00)	-	-	-**
	Dummy	1.80 (0.04)	[1.72;1.88]	44.96	<0.01
	High SES	0.15 (0.05)	[0.04;0.25]	2.69	0.01
	Low SES	0.00 (0.05)	[-0.10;0.10]	0.02	0.98
	Dummy*High SES	-0.05 (0.06)	[-0.18;0.08]	-0.78	0.44
	Dummy*Low SES	-0.13 (0.09)	[-0.31;0.05]	-1.40	0.16

Note. *SE* = Standard error. ***The 95% confidence interval, T-value en p-value are not reported, as the inclusion of a random effect makes it difficult to calculate exact statistics (e.g., there is no agreed consensus yet on how to calculate the number of degrees and number of parameters, see e.g. Baayen, Davidson, & Bates, 2008).*

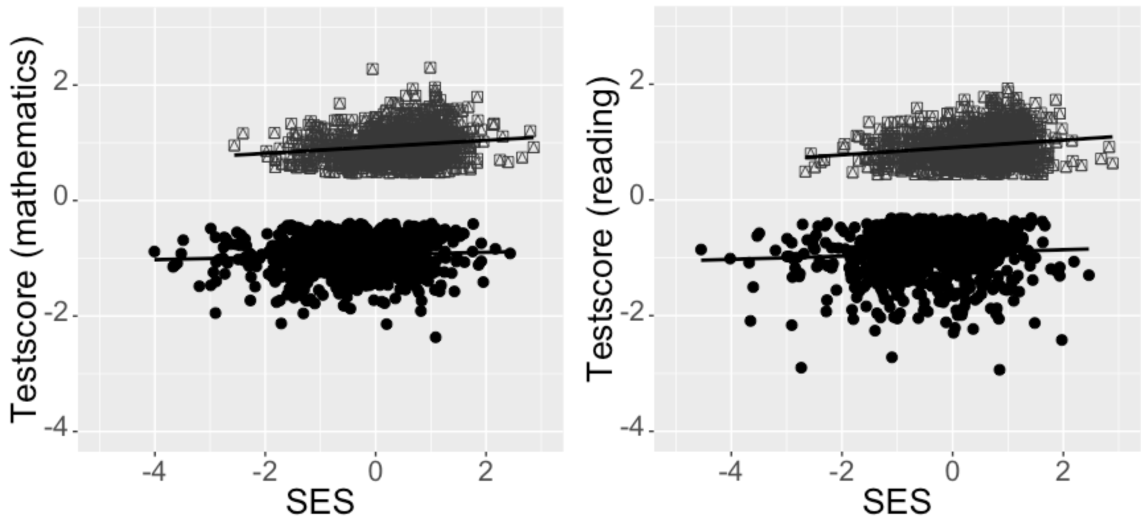

Figure 2. Effect of SES as a function of score groups (PISA sample). Left: Mathematics scores, right: mathematics scores (gray dots = 20% highest-scoring students).

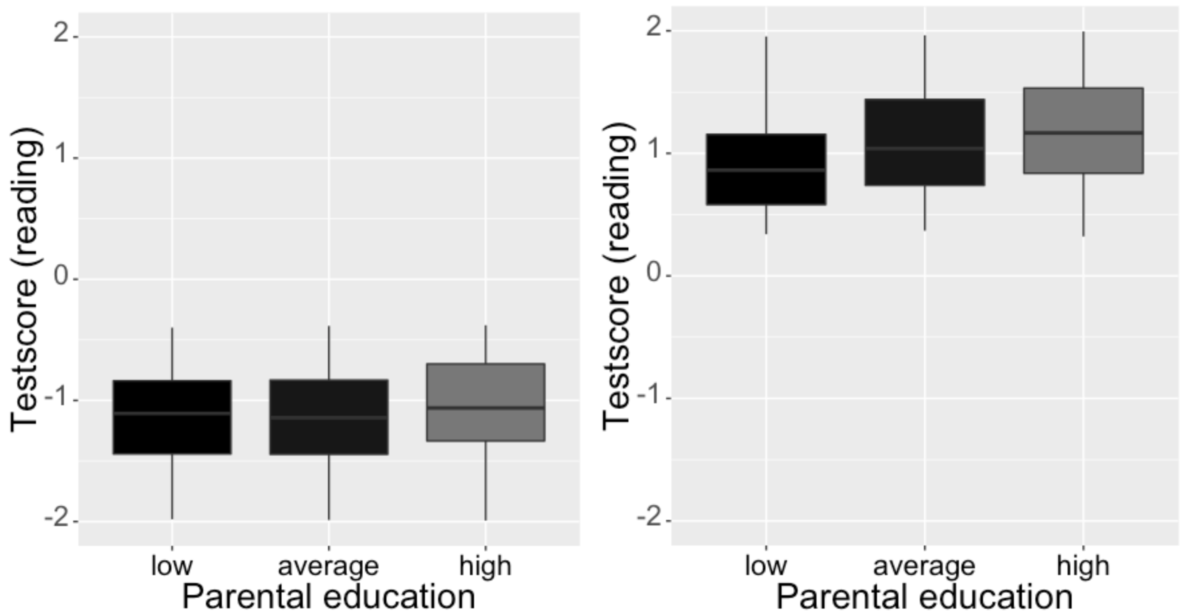

Figure 3. Effect of SES as a function of score groups (PIRLS sample). Left: low-scoring students, right: high-scoring students.

The results will be discussed separately for both domains.

Mathematics

Table 5 shows that the interaction effect between the dummy variable and individual test score is positive and significant. In Figure 2, it can be seen that the regression line for the 20% lowest-scoring students has a steeper slope than the regression line for the 20% highest-scoring students.

Reading

Table 6 shows that the interaction effect between the dummy variable and individual SES is positive and significant in the PISA sample. Figure 2 (on the right) furthermore shows that the slope of the regression line belonging to the 20% lowest-scoring students is less steep than the regression line of the 20% highest-scoring students.

As SES was measured as a categorical variable in the PIRLS sample, results of the analysis have to be interpreted in relation to a reference group (in this case, average SES). The interaction effect with the dummy variable is negative for the low-SES as well as for the high-SES group. Concerning the low-SES group, this means that the difference between the test scores of low-SES students and average-SES students is larger for high-scoring students than for low-scoring students. Concerning the high-SES group, the positive interaction effect means that the difference in scores between high-SES students and average-SES students is larger for high-scoring students than for low-scoring students. Both interaction effects are, however, non-significant (see Figure 3).

Conclusion

In this paper, we investigated the test scores of the high-scoring and low-scoring students in the domains of mathematics (PISA 2012 and TIMSS 2011) and reading (PISA 2012 and PIRLS 2011). Our first research question concerned the relationship between test scores and school influences for high-scoring and low-children respectively and was answered by means of separate multilevel analyses for both groups. The second research question investigated whether the relationship between test scores and SES was different for high-scoring and low-scoring children. This research question was answered by estimating an interaction effect between individual SES and a dummy variable that divided students into two different groups (i.e., 1 = high-scoring students, 0 = low-scoring students). In the following, results will be discussed for each research question separately.

Contrary to our expectations, the multilevel analyses showed that the between-school variance in mathematics scores was smaller for the high-scoring students than for the low-scoring students. This could be observed for both the TIMSS and the PISA sample, but was not significant. The same pattern could be observed in both samples (PISA and PIRLS) when reading scores were investigated, but was not significant either. Thus, high-scoring students are not more sensitive for school influences than low-scoring students. Note that, as we looked at between-school variance, we cannot draw any conclusion on the importance of *specific* school influences such as the quality of the teachers or the type of class (e.g., Leonardo classes) for students.

The second research question, which concerned SES influences, was answered by estimating an interaction effect between a dichotomous dummy variable (i.e., 1 = high-scoring, 0 = low-scoring) and individual SES scores. Results showed that, in the PISA sample, the relationship between SES and test score was stronger for high-scoring students than for low-scoring students. This suggests that high-scoring students are more sensitive for the influence of individual SES. This pattern could be observed for both reading and mathematics scores. However, the results of the PISA analysis (see Figure 2) also show that the variance in SES was smaller for the high-scoring children: only above a specific value of the SES variable do students show higher school performance. The same interaction effect between high-scoring performance and SES was not significant in the PIRLS sample. This may be explained by the reduced statistical power due to the use of a three-category variable of SES, which was based only on the education of the parents whereas the SES indicator provided in the PISA data was a continuous measurement based on multiple facets of SES. Furthermore, there might be bias due to missingness: as indicated, 53% of the SES values were missing for the low-scoring students.

We can conclude that there is no evidence that the influence of the specific school is more or less important for high-scoring or low-scoring students. Concerning SES, we can conclude that the relationship between SES of the parents and school performance of their child was significantly stronger for the high-scoring students than for the low-scoring students. This is surprising because SES can also be seen as a proxy for high performance:

SES is positively correlated to school performance of the parents, which, of course, pass on their talent to the next generation (Bartels et al., 2002). Evidently, the SES of the parents is not only a proxy for the academic success of their children, but has an additional positive effect for high-scoring children. In the PISA sample, we can clearly see (Figure 2) that students in the 20% lowest-performing group come from all social levels. Although the variance in SES is clearly less wide for the 20% highest-scoring children, SES was a weaker predictor of performance for the low-scoring group than for the high-scoring group.

Based on this research, it is not possible to draw any conclusion about the factors, correlated with SES, that are important in explaining why other talented children (e.g., from Asian countries) are better performing than Dutch highest performing students. Future research should continue to investigate the importance of specific factors that can positively influence the seeming underperformance of the talented students in the Netherlands.

Acknowledgements
This study was funded by the Netherlands Scientific Organization (NWO-PROO 411-12-623).

References

Baharudin, R., & Luster, T. (1998). Factors related to the Quality of the Home Environment and Children's Achievement. *Journal of Family Issues*, *19*, 357-404.
Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.
Bartels, M., Rietveld, M. J., Van Baal, G. C., & Boomsma, D. J. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Res.,5*, 544-553.
Boomsma, D.I., Vink, J., Beijsterveldt, C., de Geus, E., Beem, A., Mulder, E. et al. .. & van Baal, G. (2002). Netherlands twin register: A focus on longitudinal research. *Journal of Twin Research and Human Genetics*, *5*, 401-506.
Cigman, R. (2006). The gifted child: A conceptual enquiry. *Oxford Review of Education*, *32*(2), 197-212.
Donnelly, P., Plomin, R., & Spencer, C. C. A. (2014). The correlation between reading and mathematics at age twelve has a substantial genetic component. Nature communications 5 doi:10.1038/ncomms5204
Doolaard, S., & Oudbier, M. (2010). Het onderwijsaanbod aan hoogbegaafde leerlingen in het basisonderwijs. Utrecht: Inspectie van het Onderwijs.
Dronkers, J., & van der Velden, R. (2012). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using cross national PISA data, edited by M. Windzio, in Integration and Inequality in Educational Institutions. Dordrecht/Heidelberg/ London/New York: Springer.
Eamon, M. K. (2005). Social-demographic, School, Neighbourhood and Parenting Influences on Academic Achievement of Latino Young Adolescents. *Journal of Youth and Adolescence*, *342*, 163-175.
Ganzeboom, H. B. G., Graaf, P., de Treiman, D. J., & De Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*, 156.
Hochschild, J. L. (2003). Social Class in Public Schools. *Journal of Social Issues*, *594*, 821-840.
IBM Corp (2011). IBM SPSS Statistics for Windows, Version 20.0. Released 2011. Armonk, NY: IBM Corp.
Martin, J., Hamilton, B., Sutton, P., Ventura, S., Matthews, T., & Kirmeyer, S. (2010). Births: Final data for 2007. *National Vital Statistics Report, 58*, 1-85.
Meelissen, M. R. M., & Luyten, H. (2011). Schooleffectiviteit en prestatieniveau natuuronderwijs in groep 6: secundaire analyses op TIMSS-20007 data. *Pedagogische studiën*, *88*(5), 309-322.
Meelissen, M. R. M., Netten, A., Drent, M., Punter, R. A., Droop, M., & Verhoeven, L. (2012). PIRLS-en TIMSS-2011. Trends in leerprestaties in Lezen, Rekenen en Natuuronderwijs.Nijmegen: Radboud Universiteit, Enschede: Universiteit Twente.
Ministerie OCW (2008). Stimuleren excellentie basisonderwijs. Den Haag. Verkregen op 2202-2015 via http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/kamerstukken/2008/09/03/stimuleren-excellentie-basisonderwijs/46491.pdf
Minne, B., Rensman, M., Vroomen, B., & Webbink, D. (2007). Excellence for productivity? Bijzondere Publicatie 69, Den Haag: Centraal Planbureau.
Mooij, T., Hoogeveen, L., Driessen, G., Hell, J., & van Verhoeven, L. (2007). Succescondities voor onderwijs aan hoogbegaafde leerlingen. Eindverslag van drie deelonderzoeken. Nijmegen: Radboud Universiteit, CBO/ ITS.
Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Organisation of Economic Co-operation and Development (OECD) (2009). PISA data analysis manual. OECD Publications.
Organisation of Economic Co-operation and Development (OECD) (2003). PISA data analysis manual (SAS). OECD Publications.
Plomin, R., & Haworth, C. M. A. (2009). Genetics of high cognitive abilities. *Behavior Genetics*, *39*, 347-349.
Renzulli, J. S., Reis, S. M., & Smith, L. H. (1981). *The revolving door identification model*. Mansfield Center, CT: Creative Learning Press.
Schwabe, I., Boomsma, D. I., & van den Berg, S. M. (2017). Increased Environmental Sensitivity in High Mathematics Performance. *Learning and Individual Differences, 54,*196-201.
Shakeshaft, N., Trzaskowski, M., McMilan, A., Rimfeld, K., Krapohl, E., Haworth, C., & Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *Plos One*, 8 (*12*).
United Nations Educational, S. and C. O. (2006). International Standard Classification of Education 1997. Paris: UNESCO.
Van den Oord, E. J. C. G., & Rowe, D. C. (1998). An Examination of Genotype-Environment Interactions for Academic Achievement in an U.S. National Longitudinal Survey. *Intelligence*, *25(3)*, 205-228.
Van Houten, E. (2009). Pedagogisch quotiënt en intelligentiequotiënt. In: Minnaert, A., Lutje Spelberg, H. & Amsing, H. (Eds.), Het pedagogisch quotiënt. Pedagogische kwaliteit in opvoeding, hulpverlening, onderwijs en educatie (p. 261 t/m 282). Houten: Bohn Stafleu van Loghum.
Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large scale assessments* (*vol. 2*). IEA-ETS Research Institute.
Van der Steeg, M., Vermeer, N., & Lanser, D. (2011). Nederlandse onderwijsprestaties in perspectief. Den Haag: Centraal Planbureau.
Van Leeuwen, M., Van den Berg, S. M., & Boomsma, D.I. (2008). A twin-family study on general IQ. *Learning and Individual Differences, 18*, 76-88.
Vinkhuyzen, A. A., Van der Sluis, S., de Geus, E. J., Boomsma, D. I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes Brain Behav*, *9* (3), 276-287.
Woolfolk, A., Hughes, M., & Walkup, V. (2008). *Psychology in Education.* Harlow, England: Pearson Education Limited.

The Authors

Dr. Inga Schwabe studied psychology and did her PhD research on psychometric applications in the field of behavior genetics at the department of Research Methodology, Measurement and Data Analysis at the University of Twente (Netherlands). She worked as postdoctoral researcher at the University of Münster (Germany) and is now an assistant professor at the department of Statistics and Methodology at the Tilburg University (Netherlands).

Dr. Martina Meelissen studied educational sciences and works since 1993 as an educational researcher at the Faculty of Behavioural Sciences at the University of Twente (Netherlands). She has been involved in several, both qualitative and quantitative, national and international research projects in primary, secondary, vocational and adult education such as the Trends in Mathematics and Science Study (TIMMS-2003). She is currently the national research coordinator of TIMSS-2017.

Annemiek Punter studied educational sciences and was involved in the Dutch administration of IEA's International Computer and Information Literacy Study (ICILS) 2013 and Trends in International Mathematics and Science Study (TIMSS) 2011 and 2015. She is currently a PhD student at the department of Research Methodology, Measurement and Data Analysis at the University of Twente (Netherlands). Her PhD research is related to issues in psychometric modelling with international large scale assessment studies, with a special interest in modelling parental involvement and its relation to student achievement.

Dr. Stéphanie van den Berg studied psychology and pursued her PhD research on cognitive psychology at the University of Eindhoven (Netherlands). After having worked at the University of Utrecht (Netherlands) and the Free University of Amsterdam (Netherlands), she currently works as associate professor at the department of Research Methodology, Measurement and Data Analysis at the University of Twente (Netherlands). Her main interest lies in the application of Item-Response Theory (IRT) measurement models in the field of behaviour genetics.