**R&E-SOURCE** *https://journal.ph-noe.ac.at*
**Online Journal for Research and Education**
*Ausgabe 13, April 2020, ISSN: 2313-1640*

# Classroom matters: Mildly gifted students and their primary school performance in mathematics

*Gundula Wagner[1], Miriam Vock[2]*

***Abstract***

It is assumed that the relationship between intelligence (defined here as reasoning skills) and students' academic outcomes in mathematics is moderated by different internal and external factors on the individual student and classroom level. This study analyzes the question of whether classroom composition affects gifted students differently than average-ability students. Multilevel analyses were conducted using a sample of *N* = 333 Austrian primary school students, consisting of *n* = 51 mildly and moderately gifted students (IQ > 115), and *n* = 233 average-ability students (IQ 85 - 115). Data from *n* = 49 below-average students (IQ < 85) were taken into account at the class level. A classroom-specific effect of reasoning skills on gifted students' outcomes could be detected. There was significant variation in the slope of the reasoning skills predictor depending on the class students were in. Additional predictors on the class level were found to exert an influence on gifted students' performance. No evidence was found for cross-level effects. For the average-ability student subgroup, the context level explained a lower proportion of the variance. Moreover, the class regression lines indicated stable relationships between reasoning skills and mathematics achievement across classes.

| *Keywords:* | *Schlüsselwörter:* |
|---|---|
| compositional effects | Kompositionseffekte |
| gifted students | begabte Schüler |
| multilevel modeling | Mehrebenenanalyse |

## 1 Introduction

According to current theoretical conceptions of giftedness, gifted adolescents' progression from ability to eminence (Subotnik, Olszewski-Kubilius, & Worrell, 2011) or from giftedness to talent (Gagné, 2004) is moderated by numerous factors. These variables can be divided into factors that either limit or enhance the development of ability, and may be internal (e. g. personality traits, motivation, self-concept) or external (i. e. environmental and serendipitous factors, such as resources, opportunities, and support) (Subotnik et al., 2011). In school, important external variables include a gifted student's actual learning environment, which includes teachers, peers, and the class atmosphere (Ziegler & Stöger, 2009).

External moderating variables have been given special attention in studies on school effectiveness, and are referred to as "composition effects", or the influence exerted by the composition of the students within a given classroom or school with respect to social, ethnic, or performance-based characteristics. A common hypothesis is that the more conducive to learning a group's composition, the more academic success they will have (Baumert, Stanat, & Watermann, 2006). There is significant evidence that groups with a high proportion of students from families with low socio-economic status or from ethnic minority groups have a negative impact on the learning process of individual students within those groups (e. g., Benson & Borman, 2010; Lauder, Kounali, Robinson, & Goldstein, 2010; Palardy, 2013; Rumberger & Palardy, 2005). In contrast, groups in which there is a generally high level of achievement have a positive effect (e.g., Baumert & Schümer, 2001; Burns &

[1] Pädagogische Hochschule Niederösterreich, Department 5, Mühlgasse 67, 2500 Baden.
*Corresponding author: gundula.wagner@ph-noe.ac.at*
[2] Universität Potsdam, Professur für Unterrichts- und Interventionsforschung, Karl-Liebknecht-Str. 24 – 25, 14469 Potsdam.

Mason, 2002; Opdenakker & Van Damme, 2001), with high achieving students seeming to profit more than their average or even below average achieving classmates (Hattie, 2002; Lou, Abrami, Spence, Poulsen, Chambers, & d'Apollonia, 1996). Numerous composition effects can be explained by "group-based contagion theory" (Harris, 2010), in which students mainly benefit from advantaged peers who belong to the same subgroup. Subgroup identity can be based on different factors, such as gender, ethnicity, or ability.

Although classroom composition can be seen as an important factor in theoretical conceptions of giftedness, little research has been conducted to date on the effects of classroom compositional factors on gifted students' performance. A study by Freund, Holling, and Preckel (2007) covering gifted to below-average students identified a class-specific effect on teacher-assigned grades. However, the correlation between cognitive ability and school grades remained constant, confirming the thesis of a stable relationship. In other words: the more intelligent a student, the higher his or her academic outcomes, and the higher still his or her outcomes in a high-achieving class.

However, Reis and McCoach (2000) claim the opposite and attribute differences in performance levels to school factors. What is not yet clear is the potential moderating effect of class composition on the relationship between cognitive abilities and academic achievement, and whether this effect is the same for gifted and average-ability students. The present study examines cross-level effects of intelligence at the individual level, moderated by context variables at the class level, while controlling for non-cognitive internal factors.

# 2 Moderating Factors

## 2.1 Compositional Effects on the class level

In this section, we briefly outline the ways in which class composition may affect gifted students' achievement outcomes, referencing the considerable amount of literature that has been published in recent years describing the mechanisms behind compositional effects, including peer-related, teaching, and economic factors (Dumont, Neumann, Maaz, & Trautwein, 2013; Harker & Tymms, 2004; Harris, 2010; Verhaeghe, Vanlaar, Knipprath, De Fraine, & Van Damme, 2018).

*Peer Effects.* The term peer effects refers to social interactions among students within a group, based on the assumption that children tend to imitate the behavior of their peers. In their review of the literature on giftedness, Rost and Hanses (1994) reported less gender role-oriented play behavior among gifted children compared to average-ability children, but found only a small amount of evidence concerning the use of toys typical for boys. However, cognitive advantage may cause gifted students to critique group-specific behavior and may affect how they assess peer norms (Kasten, 2010). According to Harris (2010), group affiliation and identification are mainly affected by ethnicity and gender.

The definition of ethnicity varies in the literature. The majority of studies from German-speaking countries (which is the context of this study) measure socio-cultural background on the basis of students' family immigration history by asking for the main language spoken at home (e. g., Stanat, 2006; Tiedemann & Billmann-Mahecha, 2004; Wroblewski, 2012). Children of immigrants and immigrant children are generally disadvantaged in the German and Austrian school systems (Stanat, Rauch, & Segeritz, 2010), although this is mediated by the country of origin of the student's family. There is evidence that a high proportion of students with an immigrant background have lower academic outcomes compared to the average of students in their class. One of the explanations for this, at least in the Austrian and German school systems, is that immigrant families in Austria and Germany typically have a lower socio-economic status and less formal schooling (Henkel, Steidle, & Braukmann, 2014; Wroblewski, 2012). It has also been documented that socialization — in terms of motivation to achieve educational goals — differs based on ethnicity (Modood, 2004), as do attitudes towards scholastic education (Shah, Dwyer, & Modood, 2010), and the importance of education for social advancement (Zhou, 2005) (see also Khattab, 2015). At the same time, immigrant parents in Germany tend to have higher aspirations for their children than German parents with comparable socio-economic status (Becker & Gresch, 2016; Kristen, 2016). Less active use of the language of instruction at home is another reason why a high proportion of students with immigrant backgrounds influences a class's academic achievement (Driessen, 2002; Esser, 2006).

Effects of gender relations in the classroom on performance are also discussed internationally. It is often proposed that girls and boys be taught separately to improve school performance. However, Pahlke, Hyde, and Allison (2014) conclude in their meta-analysis that the performance differences between mono- and co-educational schools in controlled studies were trivial. At the same time, several studies have provided evidence that a high percentage of girls in a classroom positively affects primarily language, but also mathematics performance, mainly for female students (De Fraine, Van Damme, Van Landeghem, Opdenakker, & Onghena, 2003; Demanet, Vanderwegen, Vermeersch, & Van Houtte, 2013; Hoxby, 2000). Hoxby (2000) explains this finding by arguing that having more girls in class leads to: (a) higher reading skills; (b) fewer disruptions during lessons; and (c) reduced pressure for girls to behave in what are perceived socially to be typically feminine ways, thus allowing them to remain enthusiastic about mathematics and facilitating better teaching quality.

Class atmosphere as a composition variable has been relatively neglected as a subject of study, and not only with respect to gifted students. Nevertheless, school-specific class climate has been shown to be associated with academic achievement (Thapa, Cohen, Guffey, & Higgins-D´Alessandro, 2013). A positive class atmosphere has been shown to have a significant influence on mathematics performance (Tiedemann & Billmann-Mahecha, 2004). In the present paper, the class atmosphere variable is considered from a reflexive perspective—"we about us"—and is measured as the average perception of class atmosphere among a group of students.

*Teaching Effects.* How teachers instruct their students depends on those students' characteristics. Adapting one's teaching to an intermediate level of performance may reduce gifted students' opportunities to learn. With respect to class-level performance, data from several studies suggest that high-performing students tend to benefit from homogeneous learning groups (Hattie, 2002; Kim, Lee, & Lee, 2008; Lou et al., 1996; Luyten & van der Hoeven-van Doornum, 1995). These results lead to the conclusion that gifted students are likely to benefit from teachers having high expectations in high-ability classes.

In contrast, the quality of teachers' performance also depends on their approach to heterogeneity, which can be seen as a manifestation of their individualization and differentiation within the teaching process. Preckel and Vock (2013) considered such action *de rigueur* for appropriately supporting gifted students within the classroom environment. Nevertheless, in Hattie´s meta-analysis (2009), individualization in the classroom exhibited a smaller effect size on students' achievement.

Class size as a compositional factor is neither a teaching effect in the narrow sense, nor is there any empirical evidence regarding its specific relation to students' achievement. Studies claiming a positive effect of class size on academic outcomes are relatively equal in number to studies asserting the opposite, but none have investigated the effect of class size on gifted students (Ehrenberg, Brewer, Gamoran, & Willms, 2001; Hattie, 2009). However, based on the assumption that smaller classes reduce the burden on teachers and enhance teacher-student interactions, class size may affect gifted students' academic outcomes as well.

*Economic effects.* This term refers to differences in the social composition of schools, which has direct effects on their financial resources and an indirect influence on teaching materials as well as teachers' motivation. Economic resources also include increased parental involvement in schools with a socially advantaged student body (Opdenakker, Van Damme, De Frain, Van Landeghem, & Onghena, 2002). In his literature review, Khattab (2015) summarized that socio-economic disadvantage can be compensated by parental involvement and high aspirations concerning their children's performance in school. Tiedemann and Billmann-Mahecha (2004) observed an effect of parental involvement at both the individual and the class levels, which is a reasonable indicator of likely effects on gifted students' academic outcomes as well.

## 2.2 Individual Effects on the Student Level: Non-Cognitive Moderator Variables

*Self-concept* refers to individuals' mental models of their capabilities and attributes (Marsh, 1986) and is one of the key individual characteristics explaining differences between higher-achieving and underachieving gifted students (Reis & McCoach, 2000). Research shows that underachieving bright students display low self-concepts (McCoach & Siegle, 2003; Reis & McCoach, 2000). Moreover, differences in self-concept can appear as early as primary school, and the differences between gifted and non-gifted girls are greater than among boys (Loeb & Jay, 1987). Boys tend to have higher scores than girls on mathematics self-concept (Preckel, Goetz, Pekrun, & Kleine, 2008).

*Subject-specific interest* refers to a relationship between an individual and a subject that is perceived as emotionally positive and self-initiated. One result of the Study of Mathematically Precocious Youth (SMPY) was to highlight the importance of differences in interests, which seem to play an important role in performance (Lubinski & Benbow, 2000; 2006). A study by Tai, Liu, Maltese, and Fan (2006) investigating whether science-related interests predict science-related careers came to a similar result: students with average mathematics scores but science-related interests were more likely to attain a baccalaureate degree in the physical sciences or engineering than high achievers in mathematics without science-related interests.

*Gender differences* in mathematics achievement have been reported; some studies indicate slightly higher performance among male students (e. g., Leitgöb, Paseka, Bacher, & Altrichter, 2012; Wroblewski, 2012; Zimmer, Burba, & Rost, 2004). The same is true for gifted secondary school students (Preckel et al., 2008). A meta-analysis reported no differences between girls and boys (Lindberg, Hyde, Petersen, & Linn, 2010).

*Student's age.* In some classrooms, students have a broad age distribution. This may be the result of, for example, early or delayed entrance to school, grade skipping or grade retention among some students. A student's age compared to others in a given classroom therefore reflects past decisions based on that student`s ability to learn and probably degree of giftedness. A recent study by Kretschmann, Westphal, and Vock (in preparation) revealed that, contrary to former studies (e. g., Thoren, Heinig, & Brunner, 2016), the youngest students in a given class are not systematically disadvantaged. Instead, those who are younger because of acceleration (early entrance, grade skipping) are expected to exhibit higher achievement.

# 3 Research Question

While school effectiveness studies concentrate on main compositional effects only, there is a need for researchers in gifted education to learn more about external variables that affect the impact of intelligence on scholastic achievement. Assuming (a) an absence of multiple-group identification, but that students mainly benefit from advantage peers belonging to the same subgroup (Harris, 2010) and (b) that cognitive advantage may affect the way gifted students assess peer norms (Kasten, 2010), we expected different effects of class composition on gifted students' outcomes in contrast to those of average-ability students. In particular, we expected higher regression coefficients for average cognitive level, parental educational involvement, more girls in class, class atmosphere, and teaching quality among gifted students compared to average-ability students. In contrast, we expected more negative coefficients among gifted students for class size and the proportion of students with immigrant backgrounds. In accordance with moderator theories (Gagné, 2004; Subotnik et al., 2011), we assumed that class composition moderates the correlation between intelligence and performance. This is why we tested for cross-level effects between reasoning on the individual level and class composition factors on the group level. To conclude, two research questions were defined:

- Does class composition affect gifted students' achievement outcomes in the same way it does for average-ability students?
- Are there cross-level interactions between reasoning skills on the individual level and compositional factors on the class level?

Thus, we conducted an exploratory study with the purpose of specifying a model of giftedness that takes hierarchical data into consideration.

# 4 Methods

## 4.1 Data collection

To ensure a representative stratification with respect to students' social backgrounds, participating classes were selected across the five income brackets used to measure social status in the city of Vienna (Statistik Austria, 2013). One class per 100,000 inhabitants was selected within each income bracket. Informed consent from the classroom teacher and the school principal was obtained. Additionally, a written declaration of consent had to be obtained from the city school district (Stadtschulrat). The final stratified sample consisted of 18 classes: 3 classes from the zones in the highest income bracket, 7 classes from the zones in the lowest income bracket, and 8 classes from the zones in income brackets between the highest and the lowest.

Data collection was conducted in the spring of 2015. All students in a given class took part in testing as long as their parents provided written consent for their participation (81.5 %). Data collection for each class took place in a single session conducted by one of the authors and took about 100 min, including instructions and breaks. Students completed the nonverbal reasoning scales, the mathematics achievement test, and the student questionnaire, including the scales for class and teaching atmosphere, mathematics self-concept, and mathematics interest, in this order. All the students were so-called "regular" students; i. e., those whose German language competence was sufficient to understand classroom instruction. The amount of educational guidance provided by the students' parents was assessed by the teachers.

## 4.2 Sample

The *stratified* sample involved 333 primary school students from 18 3rd grade classes in 15 public and 2 private schools. All classes used the same mathematics curriculum. *The average age of the students was 9.35 years* (SD = 0.52; range = 8.00 to 11.9). The girl/boy ratio was 52.7/47.3 %. The percentage of students who were non-native German speakers in the sample was 58.3 %, which is somewhat higher than the percentage in Viennese schools as a whole during academic year 2013 – 14 (55.7 %; Statistik Austria, 2014).

*Identifying Gifted Students.* After data collection, 51 students (28 male, 23 female) in 14 classes were identified as gifted based on a cut-off score of 85 % on the nonverbal reasoning scales; thus, the classification included "mildly gifted" students (e. g., Gagné, 1993, 1998; Gross, 2000; Ziegler & Heller, 2000). In the sample as a whole, 15.3 % of students were classified as gifted (of these, 94 % were "mildly gifted" (IQ > 115), and 6 % "moderately gifted" (IQ > 130)), 68.3 % were of average ability (IQ 85 - 115), and 14.7 % were of below average ability (IQ < 85) (see Table 1).

| | Total Sample (*N* = 333) | | Gifted Students (IQ > 115) (*n* = 51) | | Average-Ability Students (IQ 85–115) (*n* = 233) | | Under Average-Ability Students (IQ < 85) (*n* = 49) | |
|---|---|---|---|---|---|---|---|---|
| | *N* | % | *n* | % | *n* | % | *n* | % |
| *Gender:* | | | | | | | | |
| Female | 176 | 52.90 | 23 | 45.10 | 132 | 56.70 | 21 | 42.90 |
| Male | 157 | 47.10 | 28 | 54.90 | 101 | 43.30 | 28 | 57.10 |
| *Immigration Status*[a]*:* | | | | | | | | |
| German language | 141 | 42.30 | 30 | 58.80 | 96 | 41.20 | 15 | 30.60 |
| other language | 192 | 57.70 | 21 | 41.20 | 137 | 58.80 | 34 | 69.40 |
| *Students population reported by districts´ income brackets:* | | | | | | | | |
| high | 44 | 13.20 | 15 | 29.40 | 28 | 12.10 | 1 | 0.02 |
| moderately | 149 | 44.70 | 26 | 50.90 | 104 | 44.40 | 19 | 38.78 |
| low | 140 | 42.10 | 10 | 19.70 | 101 | 43.50 | 29 | 59.20 |

*Note.* [a] Immigration status measured by language spoken in the family.

**Table 1: Participants demographics: *N, n* and frequencies (%)**

*Division into Two Subgroups.* In order to answer the first research question, two subgroups were created from the sample. Subgroup 1 consisted of *n* = 51 gifted students (the girl/boy ratio was 45.10/54.90 %; 41.20 % of students had an immigrant background; number of gifted students in each class *M* = 3.26; range = 2 to 7) out of a total of *n* = 258 students in 14 classes; no gifted students were identified in the remaining four classes. Hereafter, this subgroup will be designated as "gifted". Subgroup 2 consisted of *n* = 233 average-ability students (the girl/boy ratio was 56.70/43.30 %; 58.80 % of students had an immigrant background; number of average-ability students *M* = 18.50; range = 9 to 20) out of a total of *n* = 333 students in 18 classes (see Table 2). Hereafter, this subgroup will be designated as "average".

| | Level 1 | | Level 2 |
|---|---|---|---|
| | Gifted Students | Average-Ability Students | All Students in Class [a] |
| | *n* | *n* | *n* |
| Class 1[b] | -- | 12 | 13 |
| Class 2 | 4 | 14 | 19 |
| Class 3 | 2 | 11 | 13 |
| Class 4 | 3 | 9 | 12 |
| Class 5 | 2 | 10 | 15 |
| Class 6 | 3 | 13 | 19 |
| Class 7 | 4 | 16 | 20 |
| Class 8[b] | -- | 19 | 23 |
| Class 9 | 2 | 16 | 21 |
| Class 10[b] | -- | 11 | 18 |
| Class 11 | 5 | 15 | 23 |
| Class 12 | 7 | 14 | 22 |
| Class 13 | 2 | 13 | 20 |
| Class 14[b] | -- | 20 | 21 |
| Class 15 | 7 | 9 | 22 |
| Class 16 | 4 | 13 | 22 |
| Class 17 | 4 | 5 | 10 |
| Class 18 | 2 | 13 | 20 |
| total | 51 | 233 | 333 |

*Note.* [a] Includes below average-ability students.
[b] This classes are not included in the subgroup gifted. The total sum on Level 2 for the subgroup of the gifted is therefore *n* = 258.

**Table 2: Nested Sample Structure with Students Subgroups on Level 1 and all of The Students in Class on Level 2**

*Solution for Small Sample Size.* The bootstrapping simulation method, designed for parameter estimation within models, offers a solution for small sample sizes. Yung and Chan (1999) reviewed the evidence on the use of bootstrapping with small samples and concluded that no simple recommendations regarding a minimum sample size for the bootstrap method can be issued. Inasmuch as a study’s object of interest is represented by the regression coefficients, Maas and Hox (2005) suggest bootstrapping or other simulation-based methods for small random samples of 10 groups of five individuals each.

## 4.3 Study Measures

Predictor variables on the individual level are as follows:

- Student's age
- Gender
- Reasoning skills: Reasoning was assessed using two scales ("Sequence Completion" and "Matrices") [1] from the CFT 20-R (Weiß, 2006)[2]
- Self-concept in mathematics was measured using a corresponding scale from the student questionnaire for the PIRLS and TIMSS studies (Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens [BIFIE], 2011); the scale comprises 9 items, e. g., "I usually do well in mathematics," (strong agreement = 4; moderate agreement = 3; moderate disagreement = 2; and strong disagreement = 1)
- Subject-specific interest in mathematics: Students' subject-specific attitudes toward mathematics were measured using a corresponding scale from the student questionnaire for the PIRLS and TIMSS studies (BIFIE, 2011); the scale comprises 6 items, e.g., "I enjoy learning mathematics." (strong agreement = 4, moderate agreement = 3; moderate disagreement = 2; and strong disagreement = 1)
- Parental educational involvement: The educational guidance provided by the parents were assessed by teachers on a five-level scale for each child (1 = no interest to 5 = very great interest) as a predictor variable for scholastic achievement
- Immigration status: The language spoken at home was determined with a survey question asking each student for their perception ("At home, …" 1 = "I (almost) always speak German"; 2 = "I sometimes speak German and sometimes another language"; 3 = "I never speak German");

The predictor variable class context was operationalized based on the following factors:

- Homogeneity of reasoning skills: Students' individual reasoning skills were aggregated at the class level by calculating the mean in each class
- Heterogeneity of reasoning skills: The standard deviation of average reasoning skills on the class level is of additional interest as a compositional factor (e. g.; Stanat, 2006; Zimmer & Toma, 2000)
- Gender composition: Number of girls in a class divided by the total number of students
- Composition of parental educational involvement: Individual parental involvement scores were aggregated at the class level
- Immigrant ratio: Number of students who were non-native speakers of German in a class divided by the total number of students
- Class size
- School and class atmosphere were measured by means of a corresponding from the student questionnaires for the PIRLS and TIMSS studies (BIFIE, 2011) and aggregated at the class level; the scale comprises 7 items, e. g., "I feel very good in class," strong agreement = 4; moderate agreement = 3; moderate disagreement = 2; and strong disagreement = 1
- Teaching atmosphere was measured by means of the "Individualization" and "Teaching Standards" scales from the student questionnaire for the PIRLS and TIMSS studies (BIFIE, 2011) and aggregated at the class level. Exploratory factor analysis provided no evidence for the teaching standards factor; the individualization scale comprises 6 items, e. g., "The teacher would tell me whether I performed well or should continue practicing".

Mathematics achievement functioned as a response variable and was assessed with the two available sub-tests "Calculation" and "Word Problems" of the General Scholastic Achievement Test for the Third Grade (Fippinger, 1991). Due to differences between Austrian and German mathematics curricula, only Form B of the word problems scale was utilized and had to be reduced by three items[3].

| Variable | Items | *n* | *M* | *SD* | *α* | Range Potential | Range Actual | Skew |
|---|---|---|---|---|---|---|---|---|
| Self-concept | 9 | 295 | 3.18 | 0.64 | 0.85 | 1-4 | 1.0-4.0 | -0.89 |
| Subject-specific interest | 6 | 315 | 3.51 | 0.61 | 0.82 | 1-4 | 1.0-4.0 | -1.49 |
| Class atmosphere | 7 | 324 | 3.52 | 0.51 | 0.77 | 1-4 | 1.0-4.0 | -1.78 |
| Individualization in Teaching | 6 | 316 | 2.77 | 0.59 | 0.65 | 1-4 | 1.0-4.0 | -0.10 |
| Reasoning | | | | | | | | |
| Sequence Completion | 27 | 333 | 16.92 | 7.22 | 0.86 | 0-27 | 3.0-24.0 | -0.98 |
| Matrices | 27 | 333 | 17.57 | 6.04 | 0.89 | 0-27 | 4.0-25.0 | -1.23 |
| Mathematical achievement | | | | | | | | |
| Calculation | 11 | 333 | 7.67 | 2.81 | 0.83 | 0-11 | 1.0-11.0 | -0.50 |
| Word Problems | 18 | 333 | 8.77 | 3.53 | 0.79 | 0-18 | 1.0-15.0 | 0.12 |

*Note*. The variation in sample size is due to missing data.

**Table 3: Psychometric Properties of the Student Questionaires**

## 4.4 Data Analysis

Before analyzing the data, all variables were checked for statistical outliers and a normal distribution. The age variable had a left-tailed distribution, which was converted into a normal distribution by means of square transformation. The self-concept, interest, reasoning, mathematics achievement and class atmosphere variables had right-tailed distributions and were converted by means of reciprocal transformation (Reisinger, Svecnik, & Schwetz, 2012). The mathematics performance data, as the dependent variable, were further transformed into a T-metric ($M$ = 50; $SD$ = 10) to avoid negative values after reciprocal transformation. Predictor variables were z-standardized to allow for comparability of effects. Missing values were replaced using the MCMC multiple imputation technique in the SPSS program (version 22). Analyses were based on the average values obtained from five imputed data sets.

The first research question concerned potentially different effects of class composition among gifted students in contrast to average-ability students. Answering this question required estimating the coefficients separately for each group (Harris, 2010). Discrete modeling was conducted to evaluate the cohort-specific assessment pattern. Because this study focused on differences in context effects on scholastic achievement among the adjacent groups of gifted and average-ability students, effects for the subgroup of below average-ability students were not tested.

The second research question was addressed using hierarchical linear modelling. Modeling was conducted on two levels, with students on Level 1 and classes on Level 2. Using the lmer function in the “lme4” package for *R* (Bates, Maechler, Bolker, & Walker, 2015; Bates et al., 2017), maximum likelihood estimates (MLE) of the parameters of these linear mixed-effects models were determined. Following the exploratory bottom-up procedure introduced by Hox (2010), the analyses began with an empty Random-Intercept-only-Model and proceeded to add the parameters on the student and class levels.

As a first step, a Random-Intercept-only-Model (Model 1) was analyzed to provide a benchmark value for the deviance, indicating how well the model fit the data. The model was specified using an lmer call with a formula including both fixed- and random-effects terms. After specifying the Random-Intercept-only-Model, all explanatory variables at the student level were included in Model 2. Hence, this model explains mathematics performance based on the fixed effects of reasoning skills, age, gender, self-concept, interest, parents’ educational involvement and immigration status. As mentioned above, the predictor variables were z-standardized via grand mean centering for better interpretation of the results (Hox, 2010). The means and standard deviations of these intercepts were parameters to be estimated. As Model 2 only included variables on student level, only intercepts at this level were estimated.

In Model 3, all of the variables at the class level were added to the variables on Level 1 in order to evaluate whether these class-level variables explain between-group variation in the response variable. The specification of the random part of the model remained identical to the previous model, as the effect of the predictors was specified to not vary between groups. Model 3 is known as a random intercept model, as all of the class regression lines have the same slope; only the intercepts can vary.

Model 4 made it possible to illustrate the differential effect of a predictor within a given context. In this random slope model, we allowed the reasoning predictor to vary between groups. In other words, the effect of reasoning skills on math test scores can vary between classes. According to Hox (2010), a predictor variable with no significant average regression slope may have a significant variance component for this slope. Finally, Model 5 investigated cross-level interactions between the reasoning predictor and predictors at the class level.

## 4.5 Results

### 4.5.1 Research Question 1

The first research question concerned potentially different effects of class composition for the subgroups of gifted and average students. A first piece of evidence was obtained by calculating the Intraclass Correlation Coefficient (ICC) of variance components in a one-way analysis of variance (ANOVA) utilizing the "ICC" package for the statistics program *R* (Wolak, 2016). An ICC of $\rho = .11$ was obtained for the group of gifted students, indicating that 11 % of the observed differences in mathematics performance can be attributed to differences *between classes,* whereas 89 % stem from differences *among individual students within each class*. For the group of average-ability students, a lower ICC of $\rho = .06$ was detected, indicating that 6 % of the observed differences in mathematics performance can be attributed to differences *between classes,* whereas 94 % stem from differences *among individual students within each class*. According to Peugh (2010), multi-level modeling is appropriate for $\rho > 0$.

When comparing the models to one another, we observed a difference in goodness-of-fit between the models for the two subgroups. In general, models with a lower deviance fit better than models with a higher deviance. If deviance increases from one model to the next, the former model can be seen as more appropriate (Hox, 2010). In the present study, the consistent Akaike´s information criterion (CAIC) was used as a benchmark, which is preferable in complex models (Hox, 2010). Comparing the models using the "MuMln" package for *R* (Barton, 2016), the deviance for the average-ability group indicated a good model fit for Model 2, while Model 4 yielded the best fit for the gifted group. These statistical findings indicate that the final model for the average-ability group is Model 2, while it is Model 4 for the gifted group. However, in order to address to the first research question, the results of Model 2 for the average and gifted groups will be compared.

*Subgroup of average-ability students.* In this model, the predicted test performance of an average-ability female student with German as her native language, with all other Level 1 variables (reasoning, age, gender, self-concept, interest, parental involvement, and immigration status) held constant at their means, was 48.54 points. The value of the reasoning predictor (Beta = 4.72) represents the difference in test performance due to cognitive skills, after controlling for all other variables in the model. This indicates that after controlling for these variables, more intelligent students are expected to score 4.72 points higher than less intelligent students. The effect of the gender predictor is 3.78, which means that after controlling for all other variables in the model, boys outperformed girls by 3.78 points. The self-concept (3.13 points) and parental involvement (1.30 points) predictors also exhibited significant effects on test performance. No effects were detected for age, interest or immigration status.

Nakagawa and Schielzeth's (2013) pseudo-$R^2$ served as a goodness-of-fit measure for the overall model. Comparable to $R^2$ in regression analysis, it is a measure of what proportion of the variance in the response variable is explained by a specific model. For Model 2, values for the marginal $R^2$ for the proportion of variance explained by the fixed factors (0.36) as well as the conditional $R^2$ for the proportion of variance explained by the fixed and random factors together (0.38) were calculated. The "sjstats" package for *R* (Lüdecke, 2017) allows ICCs to be computed separately for each level. The ICC for Level 2, which can be interpreted as the residual variance on Level 2, was $\rho = .03$. For the average-ability subgroup, 97% of the final model was explained by within-class variance (52.83 points), and 3 % by between-class variance in the intercepts (1.72 points).

*Subgroup of gifted students.* The mean mathematics test score in Model 2 was Beta = 52.27 for a gifted girl with German as her native language. When controlling for all other variables in the model, gifted boys outperformed their female peers by 4.69 points. Students with a high self-concept outperformed students with a low self-concept by 3.48 points. Just as for the group of average students, no effects were detected for age, interest or immigration status. In addition, parental involvement did not exhibit a significant effect in this subgroup. The marginal $R^2$ and the conditional $R^2$ explained 39 % of the variance explained by the fixed and random factors.

The first research question can therefore be answered as follows: While reasoning was a significant predictor for the average-ability group in Model 2, it was not for the gifted group. The same was true for parental involvement. However, gender and self-concept explained mathematics test performance in both subgroups, although the coefficients differed.

| Parameter | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Gifted | Average | Gifted | Average | Gifted | Average | Gifted | Average |
| | | | | Fixed effects | | | | |
| Intercept | 58.36 (1.01) | 49.87 (0.81) | 52.27 (3.11) | 48.54 (0.89) | 55.04 (3.19) | -- -- | 54.57 (4.28) | -- -- |
| Level 1 (student –specific) | | | | | | | | |
| Reasoning | | | 1.53 (1.97) | 4.72* (0.94) | 3.38 (1.73) | -- -- | 4.26 (2.71) | -- -- |
| Age | | | -1.09 (0.69) | -0.68 (0.61) | 0.92 (1.12) | -- -- | 1.36 (1.03) | -- -- |
| Gender[a] | | | 4.69* (1.87) | 3.78* (1.03) | 6.00* (1.71) | -- -- | 5.78* (1.54) | -- -- |
| Self-concept | | | 3.48* (1.25) | 3.13* (0.56) | 3.79* (1.22) | -- -- | 3.64* (1.09) | -- -- |
| Subject specific interest | | | -1.57 (1.03) | -0.65 (0.54) | -1.19 (0.87) | -- -- | -0.88 (0.80) | -- -- |
| Parental involvement | | | -0.40 (1.26) | 1.30* (0.51) | -4.91* (1.44) | -- -- | -5.72* (1.29) | -- -- |
| Immigration status[b] | | | -1.54 (1.96) | -0.47 (1.01) | -1.18 (1.77) | -- -- | -0.91 (1.74) | -- -- |
| Level 2 (class) | | | | | | -- -- | | -- -- |
| Homogenity in reasoning skills | | | | | -4.14 (3.14) | -- -- | -4.25 (2.84) | -- -- |
| Heterogenity in reasoning skills | | | | | -5.24* (1.98) | -- -- | -5.84* (1.76) | -- -- |
| Class atmosphere | | | | | 1.90 (1.43) | -- -- | 2.25 (1.26) | -- -- |
| Individualization in teaching | | | | | 0.56 (1.59) | -- -- | 0.77 (1.41) | -- -- |
| Class size | | | | | -6.37* (1.78) | -- -- | -7.03* (1.59) | -- -- |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Composition of parental involvement | | | | | | | | | -2.60 | (2.60) | -- | -- | -2.49 | (2.30) | -- | -- |
| Immigration ratio | | | | | | | | | -7.79* | (3.41) | -- | -- | -7.88* | (3.05) | -- | -- |
| Gender composition | | | | | | | | | -2.70* | (1.00) | -- | -- | -3.22* | (0.94) | -- | -- |
| Marginal $R^2$ explained by the fixed paramters | | | | | 0.39 | | 0.36 | | 0.61 | | -- | | 0.57 | | -- | |
| | | | | | | | Random parameters | | | | | | | | | |
| Level 2 | | | | | | | | | | | | | | | | |
| Intercept $\tau_{00}$ | 0.00 | (0.00) | 5.30* | (2.30) | 0.00 | (0.00) | 1.72* | (1.31) | 0.00 | (0.00) | -- | | 83.90* | (9.43) | -- | |
| Reasoning $\tau_{11}$ | | | | | | | | | | | | | 39.13* | (6.47) | -- | |
| Cov (Reasoning/intercept) $\tau_{01}$ | | | | | | | | | | | | | -57.12 | --[c] | -- | |
| Level 1 | | | | | | | | | | | | | | | | |
| Intercept $6^2$ | 51.6* | (7.18) | 80.54* | (8.97) | 31.64* | (5.62) | 52.83* | (7.26) | 20.43* | (4.52) | -- | | 15.07* | (3.93) | -- | |
| -2* log likelihood | -172.93 | | -847.37 | | -160.45 | | -800.03 | | -160.45 | | -795.91 | | -148.01 | | -- | |
| cAIC | 347.85 | | 1697.06 | | 336.90 | | 1610.21 | | 330.61 | | 1614.44 | | 309.87 | | -- | |
| Conditional $R^2$ explained by the fixed and random parameters | | | | | 0.39 | | 0.38 | | 0.61 | | -- | | 0.72 | | -- | |

*Note*. Standard errors are in parentheses.
[a] Females serve as a benchmark category.
[b] Students with German language serve as a benchmark category.
[c] R does not calculate the standard error.
* $p < 0.05$.

**Table 4: Fixed Effects Estimates (Top) and Variance-Covariance Estimates (Bottom) for Models of the Predictors of Mathematical Achievement**

### 4.5.2 Research Question 2

Model 5 addressed the second research question concerning cross-level effects between reasoning on Level 1 and class composition on Level 2. As discussed above, no predictors on the class level were detected at all in the average-ability group. For the gifted group, Model 5 had a weaker model fit than Model 4. Furthermore, the analysis of Model 5 revealed no significant interaction effects. Model 5 was omitted from Table 4 due to its complete lack of significant effects. Consequently, the second research question cannot be answered definitively, as Model 5 has been omitted from Table 4. Instead, Model 3 and 4 will be reported due to their partially significant results.

*Subgroup of gifted students.* As already mentioned, no residual variance between classes could be detected in Model 2. However, further multilevel modeling was advised because of $\rho > 0$ in the ANOVA model. The predictors at the class level were gradually introduced into Model 3. In this model, the overall intercept (Beta = 55.04) now represents the predicted mathematics score for a gifted girl with German as her native language who scores at the mean on the mathematics test and is in a class that scores at the class mean on the test and has average values for all predictor variables on Level 1 and Level 2. We will first consider the effects of the Level 2 variables on the intercept. Heterogeneity in reasoning skills had a significant negative effect on gifted students' test scores, decreasing them by 5.24 points after controlling for all other variables in the model. This coefficient is negative — as are all other significant variables on Level 2 — indicating that, holding classroom heterogeneity in reasoning skills constant at the mean, every one-point increase in heterogeneity in a class causes gifted students' predicted mathematics performance level of gifted students to decrease by 5.24 points. All other significant variables can be interpreted in the same way; in classes with a higher ratio of girls to boys, gifted students' scores dropped 2.70 points. Class size negatively affected performance by 6.37 points. However, a high percentage of students with immigrant backgrounds had the greatest effect, translating into a reduction of 7.79 points compared to the average predicted test performance.

We next examined the variables on Level 1, controlling for all other variables in the model. Again, gifted boys outperformed girls by 6.00 points. Students with a high self-concept outperformed students with a low self-concept by 3.79 points. Moreover, parental educational involvement also exerted a considerable influence on mathematics performance at the individual level (– 4.90 points), whereas reasoning skills again exhibited no significant effect.

Nakagawa and Schielzeth's (2013) marginal $R^2$ for the proportion of variance explained by fixed factors as well as the conditional $R^2$ for the proportion of variance explained by fixed and random factors was 0.61; thus, the explained variance increased from Model 1 to Model 2. Again, no residual variance between classes could be determined, but further multilevel modeling was advised by $\rho > 0$ in the ANOVA Model.

Finally, a random-slope model (Model 4) was estimated to assess whether the slope of the reasoning predict has a significant variance component between classes. The overall intercept, which was about Beta = 54.57 points, can be interpreted as described above. Controlling for all other variables in the model, the same Level 2 variables exerted a negative influence on test performance as Model 3: heterogeneity at the cognitive level (–5.84 points), a high percentage of girls (–3.22 points), class size (–7.03 points) and, most notably, a high percentage of students with immigrant backgrounds (–7.88 points). On Level 1, boys' performance was 5.78 points higher. A high self-concept pushed gifted students' test performance up by 3.64 points, whereas a low parental educational involvement led to a decrease of 5.72 points.

Residual variance on Level 2 could be detected by taking the variance in the slope of the reasoning predictor into consideration. The "sjstats" package for *R* (Lüdecke, 2017) computed $\rho = 0.85$[4]. This means that 85% of mathematics achievement was due to the class students were in. Simultaneously, Nakagawa and Schielzeth's (2013) conditional $R^2$ for the proportion of variance explained by fixed and random factors increased to 72 %, while the marginal $R^2$ explained only by fixed factors decreased to 57 %. In this final random slope model, 10.4 % of gifted students' mathematics achievement (15.46 points) was explained by within-class variance, 61 % by between-class variance in the intercepts (88.94 points), and 28.6 % by between-class variance in the slope of the reasoning predictor (41.86 points).

### 4.5.3 Bootstrapping

Multi-level bootstrapping was conducted to improve the parameter estimations for the final models for each subgroup. A large number of iterations (b > 5000) (Hox, 2010) is required to accurately set confidence intervals. Thus, 10,000 iterations were conducted for Model 1 among the average-ability group and Model 3 among the gifted group. One of the basic requirements of multi-level bootstrapping is correctly simulating (Hox, 2010) the hierarchical data structure, which is rendered possible by the "boot" package for the statistics program *R*

(Canty & Ripley, 2016). The regression coefficients of the estimated models generally corresponded to the initial models and lay within the confidence intervals, which were checked for every coefficient (see Appendix).

## 5 Discussion

The path from giftedness to academic outcomes is believed to be moderated by sets of internal and external variables that affect students' trajectories. One set of external variables studied in educational psychology refer to the social and intellectual composition of the classrooms in which students are located (Dumont et al., 2013). In this study, we analyzed whether classroom composition affects gifted students' achievement outcomes in the same way as for average-ability students.

Specifically, we modeled the correlation between reasoning skills and mathematics performance as a function of classroom composition above and beyond non-cognitive internal factors for a group of gifted students and a group of average students. The results support the assumption that gifted students are affected by class composition differently than their average-ability peers.

For the subgroup of gifted students, a class-specific effect of reasoning skills on mathematics achievement was detected. In accordance with theoretical conceptions of giftedness (e. g., Gagné, 2004; Subotnik et al., 2011), the correlation between reasoning and mathematics achievement was moderated by external factors. Thus, the effect of an individual gifted student's intelligence depends on class composition. What remains unclear, however, is which factor(s) in particular caused the class-specific effect. Contrary to expectations, none of the observed class-level predictors moderated the effect of reasoning on mathematics achievement, as no cross-level effects were detected. There was merely evidence of an unspecific "dynamic interaction" (Gagné, 2004, p. 121) between reasoning skills on the individual level and the class context. Reasoning had a higher or lower predictive effect on test scores depending on the class, as reflected in rising or falling class regression lines (see Figure 1). The notion of a stable correlation between intelligence and outcomes, as supported by Freund et al. (2007), does not apply to gifted students in this sample. Additional predictors on the class level were found to be related to gifted students' performance. Gifted students achieved higher test scores in classes with a lower proportion of immigrant students, a lower proportion of girls, a smaller class size, and less heterogeneity in cognitive abilities.

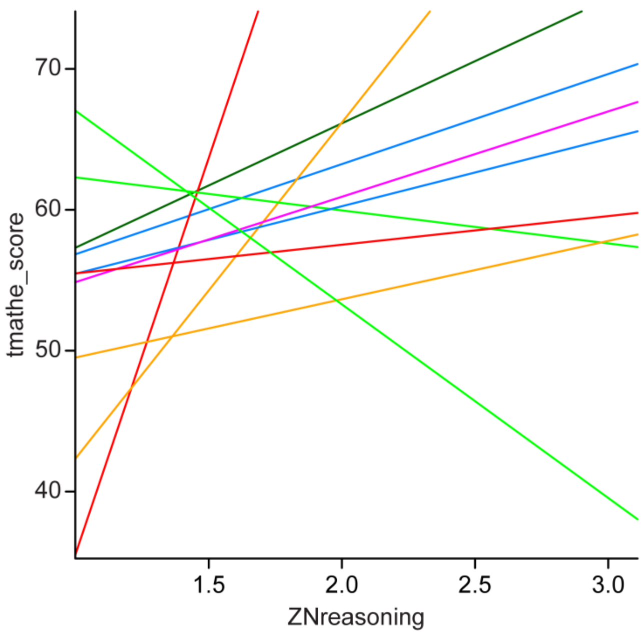


**Figure 1: The class regression lines of the individual predictor reasoning (x-axis) shown in correlation with mathematic achievement (y-axis). There is a large amount of variability between classes in mathematical test scores. A steeper slope of the predictor reasoning leads to higher test scores, and a dipping regression line leads to low test performance. Because of overlapping class regression lines only ten lines are visible.**

For the subgroup of average-ability students, the class regression lines had equal slopes, indicating constant relationships between reasoning skills and mathematics achievement across classrooms. The more intelligent students were, the higher their academic outcomes. Moreover, a smaller proportion of variance was explained by the context level. There were only small differences between classes in terms of students' success (see Figure 2). Thus, which class average-ability students were in mattered much less for their mathematics achievement. Predictors on the individual level provided a much better explanation of variance. This suggests that the variance on the class level resulted from a systematic rather than a random distribution of high-achieving students among schools and classes.

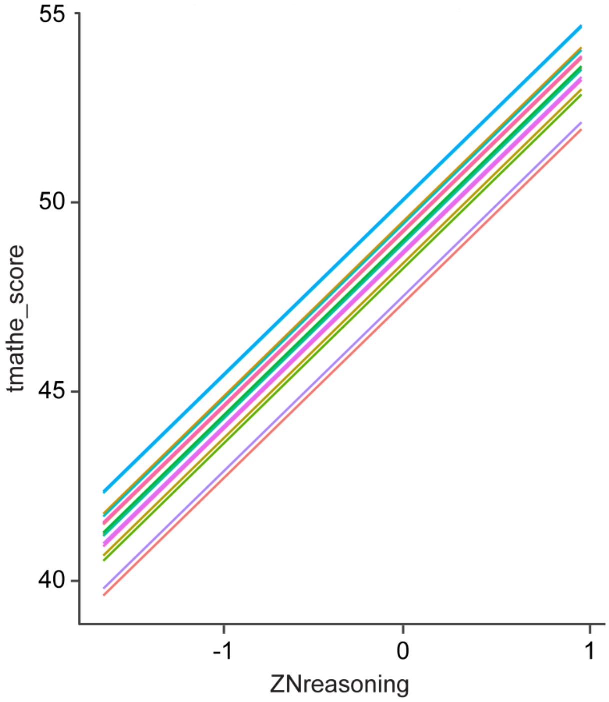


**Figure 2: The class regression lines of the individual predictor reasoning (x-axis) shown in correlation with mathematic achievement (y-axis). The class regression lines have equal slopes, which indicate stable relations between intelligence and mathematical achievement. The higher reasoning skills, the higher mathematical test scores.**

Considering the results for gifted students' academic outcomes in detail, the proportion of students with immigrant backgrounds was the strongest predictor on the class level: the more such students were in a class, the lower gifted students' test scores. However, these data must be interpreted with caution, because adding in average socio-economic status typically reveals a strongly diminishing effect of ethnic composition (e. g., Rumberger & Palardy, 2005; Van der Slik, Driessen, & De Bot, 2006). Unfortunately, in this study we were not able to assess students´ socio-economic or ethnic status. By asking about the language spoken at home, we probably indirectly addressed the effect of language skills in the language of instruction on the individual level and the effect of language ability grouping on the class level, rather than the effect of immigration status per se. If so, these results may reflect a compositional teaching effect in which teachers adapt their instruction to a lower average level of language skills in a class, leading to reduced learning opportunities for gifted students. However, these results were true for all gifted students, both those who spoke primarily German at home as well as those who spoke other languages. The language spoken at home did not seem to be related to gifted students' achievement as a predictor on the individual level.

Contrary to expectations, no effects of average parental involvement were detected; negative effects were limited to the individual level. A possible explanation may stem from the fact that this variable was assessed by teachers, who may have overestimated parental involvement among gifted students. Rather than recognizing gifted students' ability, teachers may have primarily attributed these students´ achievement to the support of their parents, which would explain the negative effect.

In contrast to earlier findings suggesting that high-achieving students' performance is not affected by classmates with lower ability levels (Dar & Resh, 1986; Resh & Dar, 1992), these results indicate that a wide distribution of student ability levels in a classroom is negatively related to gifted students' outcomes. Again, one possible explanation is a compositional teaching effect: if teachers adapt their lessons to students with lower cognitive abilities, gifted students might get bored. In turn, it is plausible to assume that gifted students who are not challenged by advanced tasks are at risk of becoming underachievers (Eddles-Hirsch, Vialle, Rogers, & McCormick, 2010; Emerick, 1992; Gronostaj, Werner, Bochow, & Vock, 2016); however, research on this relationship is still scarce (White, Graham, & Blaas, 2018). Another possible explanation concerns a peer effect mechanism based on peer group norms, with disadvantaged students pulling down the achievement of all other students in a class by inducing peer norms focused on low achievement (Hoxby & Weingarten, 2005; Jencks & Mayer, 1990). However, this interpretation runs contrary to “group-based contagion theory” (Harris, 2010), which views the notion of an overall peer-group pressure for conformity with skepticism.

Class size was negatively related with gifted students’ mathematics achievement in our analysis. Typically, the effects of class size on student outcomes are rather small (Hattie, 2009), and the possible effects of class size are controversial (Brühwiler & Helmke, 2018; Kunter & Trautwein, 2013). In previous studies, positive effects were detected specifically among socially disadvantaged children and students with lower achievement (Blatchford, Bassett, Goldstein, & Martin, 2003; Blatchford, Russell, Bassett, Brown, & Martin, 2007). The more students in a class, the less time the teacher can invest in dealing with the unique needs of each individual child. Hence, in a larger class, the teacher might invest his or her limited resources in students with learning difficulties rather than gifted children who are able to learn quite well. Our results indicate a specific effect of class size for the teaching of gifted students, similar to other groups with special needs, so that gifted students also seem to benefit from smaller classes.

One surprising result was the negative effect of a high proportion of girls on mathematics achievement. This result is difficult to explain, because the literature only has evidence of positive effects: higher in the case of language performance, lower in the case of mathematics performance (De Fraine et al., 2003; Demanet et al., 2013; Hoxby, 2000). From the perspective of a peer effect mechanism, a high percentage of girls may cause less competition in class, which is negatively related to achievement (Baker, Bridger, & Evans, 1998). This is because girls are considered more socially aware and experience different achievement pressure than boys. Accordingly, this result is likely to be related to the lack of effect of class atmosphere: the better the cohesion among the students in a class, the less competition can be assumed. According to this data, gifted students seemed to need competition in class as an incentive to perform. Again, a teaching effect mechanism is also possible. Matheis, Kronborg, Schmitt, and Preckel (2017) found that German teachers in training exhibit the greatest enthusiasm when teaching girls with average ability, which could account for the disadvantage among gifted students in the sample.

As Marsh and Martin (2011) would have us expect, academic achievement was related to the non-cognitive individual-level factor of academic self-concept in both subgroups. Thus, a reciprocal effect between academic self-concept and academic performance must be assumed. Our finding that boys outperformed girls is not supported by a recent meta-analysis (Lindberg et al., 2010). Nevertheless, the girls in this sample faced specific challenges regarding mathematics achievement in the classroom setting.

## 6 Limitations

The strength of this study is its use of an unselective sample of gifted students within general classroom settings; however, this resulted in a small proportion of moderately and higher gifted students. Our intention was to examine gifted students in regular schools, as few gifted students actually benefit from support programs tailored to their unique abilities, such as gifted classes. By not preselecting such classes, the number of gifted students in the sample was left to chance. Nevertheless, a useful normal distribution of cognitive abilities was obtained. However, the results are limited to “mildly gifted” students within their classroom settings.

For the data collection we needed the consent of the teachers and headmasters before classes could participate in the study. As this ad-hoc sample was based on voluntary participation, sampling bias cannot be ruled out. Variance on the class level could have been influenced by unobserved characteristics of the participating classes. The absence of cross-level effects between cognitive skills at the individual level and class composition parameters on the classroom level can most likely be explained by the small sample size. Maas and Hox (2005) recommend a sample size of $n$ = 50 at the aggregate level to detect cross-level effects. In general, the small sample size must be taken into account when interpreting the data, even though the results were secured by bootstrapping.

Students' socio-economic and immigration status could not be measured directly. Instead, we used available indicators for these concepts. In our study, parents' educational involvement did not mitigate the effect of the immigration ratio on school performance, as had been observed in other studies (e. g., Rumberger & Palardy, 2005; Van der Slik, Driessen, & De Bot, 2006). Additionally, teachers may have overestimated the parental educational involvement of gifted students; thus, a parent questionnaire would have been a more appropriate information source for the study. A parent questionnaire would have also made it possible to determine parents' educational background and professional status. It cannot be ruled out that measuring children's immigration status by asking about the language spoken at home may have instead measured language skills. Given the large proportion of non-native German-speaking students in the population, an additional language test would have been preferable.

Teaching atmosphere was measured only through the students' perspective. Because these students were young, they had limited insight into pedagogical issues. A more robust study would have measured teacher expertise and its impact on students' mathematics achievement (e. g., Hattie, 2009; Seidel & Shavelson, 2007). Although it is certainly preferable to assess giftedness using a variety of different measures (Worrell, 2009), in this study it was only possible to use a relatively economical assessment focusing solely on reasoning. Therefore, the definition of giftedness in this study is narrow and does not incorporate different types of giftedness. For practical reasons, we used a group speed test to measure reasoning. By doing so, individual differences in students´ intelligence could not be taken into account. However, Lavergne and Vigneau (1997) pointed out that children's mental speed strongly develops between ages 9 and 11, which corresponds to the age group of the students in the sample. The study also focused exclusively on mathematics achievement. The compositional effects among gifted students analyzed in this study should also be investigated in other domains, such as science or verbal subjects.

# 7 Conclusion

The present study has demonstrated that the classroom context is of special interest for gifted education in primary schools in Austria. However, these results have limited applicability to other student populations. Considerably more work will need to be done to determine the full extent of classroom-specific variation in the correlation between reasoning skills and mathematics performance among gifted students. Larger samples could provide more definitive evidence and are more likely to include more highly-gifted students as well. Furthermore, longitudinal studies can provide more conclusive produce evidence of causal relationships. Average prior achievement in class, for example, is one of the best predictors of further achievement.
Moreover, further research should be limited to context variables that are empirically well-supported, such as socio-economic, socio-cultural and performance-related factors. The latter should include a greater focus on composition effects concerning language skills in gifted education. This could be a fruitful area for further work and could have implications for education in the gifted classroom and beyond[5].

## *References*

Baker, J. A., Bridger, R., & Evans, K. (1998). Models of underachievement among gifted preadolescents: The role of personal, family, and school factors. *Gifted Child Quarterly, 42*, 5–15. doi: 10.1177/001698629804200102

Bates, D., Maechler, M., Bolker, B., & Walker, St. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. Doi:10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., Walker, St., Christensen, R.H.B., Singmann, H., . . . Green, P. (2017). *Linear Mixed-Effects Models using 'Eigen' and S4*. Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

Barton, K. (2016). *Multi-Model Inference.* Retrieved from https://ftp.cn.debian.org/CRAN/web/packages/MuMIn/MuMIn.pdf

Baumert, J., & Schümer, G. (2001). Schulformen als selektionsbedingte Lernmilieus [School types as selection-dependent learning environments]. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, . . . M. Weiß (Eds.), *Pisa 2000. Die Länder der Bundesrepublik Deutschland im Vergleich* [Pisa 2000. The States of the Federal Republic of Germany Compared] (pp. 454–467). Opladen, Germany: Leske & Budrich.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [School structure and the emergence of differential learning and development environments]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000* [Disparities of origin in education. In-depth analyses within the framework of PISA 2000] (pp. 95–188). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Becker, B. & Gresch, C. (2016). Bildungsaspirationen in Familien mit Migrationshintergrund [Educational aspirations in families with a migration background]. In C. Diehl, C. Hunkler, & C. Kristen (Eds.), Ethnische Ungleichheiten im Bildungsverlauf [Ethnic inequalities in education] (pp. 73–115). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Benson, J., & Borman, G. D. (2010). Family, neighborhood, and school settings across seasons: When do socioeconomic context and racial composition matter for the reading achievement growth of young children? *Teachers College Record*, *112*, 1338–1390. doi:10.1177/0013124510392779

Blatchford, P., Bassett, P., Goldstein, H., & Martin. C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the Institute of Education Class Size Study of Children Aged 5–7 Years. *British Educational Research Journal, 29,* 709–730. doi:10.1080/0141192032000133668

Blatchford, P., Russell, A., Bassett, P., Brown, P., & Martin, C. (2007). The effect of class size on the teaching of pupils aged 7–11 years. *School Effectiveness and School Improvement, 18,* 147–172. doi:10.1080/09243450601058675

Brühwiler, C., & Helmke, A. (2018). Determinanten der Schulleistung [Determinants of school performance]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), Handwörterbuch Pädagogische Psychologie [Hand Dictionary Educational Psychology] (Vol., 5, pp. 78–92). Weinheim, Germany: Beltz Psychologie Verlagsunion.

Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE). (2011). PIRLS und. TIMSS. Schülerfragebogen [PIRLS and TIMSS 20011. Student questionaires]. Retrieved from https://www.bifie.at/system/files/dl/Sch%C3%BClerfragebogen.pdf

Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. *American Educational Research Journal*, *39*, 207–233. doi:10.3102/00028312039001207

Canty, A., & Ripley, B. (2016). *Bootstrap Functions (Originally by Angelo Canty for S).* Retrieved from https://cran.r-project.org/web/packages/boot/boot.pdf

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal, 23*, 357–374. doi:10.3102/00028312023003357

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal*, *29*, 841–859. doi:10.1080/0141192032000137330

Demanet, J., Vanderwegen, P., Vermeersch, H., & Van Houtte, M. (2013). Unravelling gender composition effects on rule-breaking at school: A focus on study attitudes. *Gender and Education*, *25*, 466–485. doi:10.1080/09540253.2013.772567

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, *28*, 347–368. doi:10.1016/S0191-491X(02)00043-3

Dumont, H., Neumann, M., Maaz, K., & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde [Class composition as a defining parameter of student achievement. International and national findings]. *Psychologie in Erziehung und Unterricht*, *60,* 163–183. doi:10.2378/peu2013.art14d

Eddles-Hirsch, K., Vialle, W., Rogers, K. B., & McCormick, J. (2010). "Just challenge those high-ability learners and they´ll be all right!" The impact of social context and challenging instruction on the affective development of high-ability students. *Journal of Advanced Academics, 22*, 102–124. Retrieved from https://researchonline.nd.edu.au/cgi/viewcontent.cgi?article=1049&context=edu_article

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, *2*, 1–30. doi:10.1111/1529-1006.003

Emerick, L. J. (1992). Academic underachievement among gifted: Students ́ perception of factors that reverse the pattern. *Gifted Child Quarterly, 36,* 140–146. Retrieved from: http://www.davidsongifted.org/Search-Database/entry/A10178

Esser, H. (2006). Sprache und Integration. Die sozialen Bedingungen und Folgen des Sprachenerwerbs von Migranten [Language skills and integration. Immigrants´ language acquisition and it´s social outcomes]. Frankfurt on Main, Germany: Campus.

Fippinger, F. (1991). *Allgemeiner Schulleistungstest für die 3. Klassen* [General school achievement test for the third grade]. Weinheim, Germany: Beltz.

Freund, P. A., Holling, H., & Preckel, F. (2007). A multivariate, multilevel analysis of the relationship between cognitive abilities and scholastic achievement. *Journal of Individual Differences*, *28*, 188–197. doi:10.1027/1614-0001.28.4.188

Gagné, F. (1993). Constructs and models pertaining to exceptional human abilities. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International Handbook of Research and Development of Giftedness and Talent* (pp. 69–87). Oxford, UK.: Pergamon Press.

Gagné, F. (1998). A proposal for subcategories within gifted or talented populations. *Gifted Child Quarterly*, *42*, 87–95.

Gagné, F. (2004). Transforming gifts into talents: the DMGT as a developmental theory. *High Ability Studies, 15,* 119–147, doi:10.1080/1359813042000314682

Gronostaj, A., Werner, E., Bochow, E., & Vock, M. (2016). How to learn things at school you don´t already know: Experiences of gifted grade-skippers in Germany. Gifted Child Quarterly, 60, 31–46. doi:10.1177/0016986215609999

Gross, M. (2000). Exceptionally and profoundly gifted students: An underserved population. Understanding Our Gifted, 12, 3–9. Retrieved from http://www.hoagiesgifted.org/underserved.htm

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. School Effectiveness and School Improvement, 15, 177–199. doi:10.1076/sesi.15.2.177.30432

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. Teachers College Record, 112, 1163–1197.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449–481. doi:10.1016/S0883-0355(03)00015-6

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

Henkel, M., Steidle, H., & Braukmann, J. (2014). *Familien mit Migrationshintergrund: Analysen zur Lebenssituation, Erwerbsbeteiligung und Vereinbarkeit von Familie und Beruf* [Families with a migration background: Analyses of life situation, labour force participation and compatibility of family and career]. Berlin, Germany: Bundesministerium für Familie, Senioren, Frauen und Jugend. Retrieved from https://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/Familien-mit-Migrationshintergrund,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (Vol. 2). New York, NY: Routledge.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation* (NBER Working Paper No. 7867). Cambridge, MA: National bureau of Economic Research. Retrieved from http://www.nber.org/papers/w7867

Hoxby, C. M. & Weingarten, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi:10.1.1.472.2561&rep=rep1&type=pdf

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. E. Lynn, Jr. & Michael G.H. McGeary (Eds.), Inner city poverty in the United States (pp.111–186). Washington, DC: National Academic Press.

Jones, K. (1992). Amendment to 'Specifying and estimating multilevel models for geographical research'. Transactions of the Institute of British Geographers, 17, 246. doi: 10.2307/622549

Kasten, H. (2010). Geschlechtsunterschiede [Gender differences]._In D. Rost (Ed.), *Handbuch Pädagogische Psychologie* [Handbook of educational psychology] (Vol. 4; pp. 234–241). Weinheim, Germany: Beltz Verlag.

Khattab, N. (2015). Students' aspirations, expectations and school achievement: What really matters? *British Educational Research Journal, 41,* 731–748. doi:10.1002/berj.3171

Kim, T., Lee, J., & Lee, Y. (2008). Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea. *Economics of Education Review*, *27*, 697–711. doi:10.1016/j.econedurev.2006.07.014

Kretschmann, J., Westphal, A., & Vock, M. (in preparation). Does youth matter? Relative age effects on psychosocial adjustment, grades and teachers' secondary school recommendation.

Kristen, C. (2016). Migrationsspezifische Ungleichheiten im deutschen Hochschulbereich [Migration-specific inequalities in German higher education]. In C. Diehl, C. Hunkler & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten* [Ethnic inequalities in education. Mechanisms, findings, debates] (pp. 643–668). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts* [Psychology of teaching]. Paderborn, Germany: Verlag Ferdinand Schöningh.

Lauder, H., Kounali, D., Robinson, T., & Goldstein, H. (2010). Pupil composition and accountability: An analysis in English primary schools. *International Journal of Educational Research*, *49*, 49–68. doi:10.1016/j.ijer.2010.08.001

Lavergne, Ch., & Vigneau, F. (1997). Response speed on aptitude tests as an index of intellectual performance: A developmental perspective. *Personality and Individual Differences, 23,* 183–363. doi:10.1016/S0191-8869(97)00037-8

Leitgöb, H., Paseka, A., Bacher, J., & Altrichter, H. (2012). Geschlechtsspezifische Leistungsdifferenzen in Österreich. Vertiefende Analysen der Pisa 2009 Daten. [Gender-specific differences in performance in Austria. Detailed analysis of Pisa 2009 data]. In F. Eder (Ed.), *Pisa 2009. Nationale Zusatzanalysen für Österreich* [Pisa 2009: Additional national analyses for Austria] (pp. 337–366). Münster, Germany: Waxmann.

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

Loeb, R. C., & Jay, G. (1987). Self-concept in gifted children: Differential impact in boys and girls. *Gifted Child Quarterly*, *31*, 9–14. doi:10.1177/001698628703100102

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423–458. doi:10.3102/00346543066004423

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, *55*, 137–150. doi:10.1037/0003-066X.55.1.137

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

Lüdecke, D. (2017). sjstats: Statistical Functions for Regression Models. *R package version 0.11.0.* Retrieved from https://CRAN.R-project.org/package=sjstats.

Luyten, H., & van der Hoeven-van Doornum, A. (1995). Classroom composition and individual achievement. Effects of classroom composition and teacher goals in Dutch elementary education. *Tijdschrift voor Onderwijsresearch*, *20*, 42–62.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. doi:10.1027/1614-1881.1.3.86

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., & Martin, A.J. (2011). Academic self-concept and academic achievement: relations and causal ordering. *The British Journal of Educational Psychology, 81,* 55–77. doi:10.1348/000709910X503501

Matheis, S., Kronborg, L., Schmitt, M., & Preckel, F. (2017). Threat or challenge? Teachers' beliefs about gifted students and their relationship to teacher motivation. Gifted and Talented International, 32, 134–160. doi:10.1080/15332276.2018.1537685

McCoach, D. B., & Siegle, D. (2003). Factors that differentiate underachieving gifted students from high-achieving gifted students. Gifted Child Quarterly, 47, 144–154. doi:10.1177/001698620304700205

Modood, T. (2004). Capitals ethnic identity and educational qualifications. Cultural Trends, 13, 87–105. doi:10.1080/0954896042000267170

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4,* 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Opdenakker, M.-C., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, *27*, 407–432. doi:10.1080/01411920120071434

Opdenakker, M.-C., Van Damme, J., De Frain, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement, 13,* 399–427. doi:10.1080/0141192032000137330

Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal, 50,* 714–754. doi:10.3102/0002831213481240

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: a meta-analysis. *Psychological Bulletin, 140*, 1042–1072. doi:10.1037/a0035740

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85–112. doi:10.1016/j.jsp.2009.09.002

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students: Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted Child Quarterly*, *52*, 146–159. doi:10.1177/0016986208315834

Preckel, F., & Vock, M. (2013). *Hochbegabung. Ein Lehrbuch zu Grundlagen, Diagnostik und Fördermöglichkeiten* [A textbook on the fundamentals, the diagnosis and the fostering of giftedness]. Göttingen, Germany: Hogrefe.

Reis, S. M., & McCoach, D. B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly, 44,* 152–170. doi: 10.1177/001698620004400302

Reisinger, C.-M., Svecnik, E., & Schwetz, H. (2012). *Fehlende Werte und keine Normalverteilung? Tipps und Abhilfen für das quantitativ orientierte Forschen* [Missing values and no normal distribution? Tips and remedies for quanitative research]. Vienna, Austria: Facultas.

Resh, N., & Dar, Y. (1992). Learning segregation in junior high schools in Israel: Causes and consequences. *School Effectiveness and School Improvement, 3*, 272–292. doi:10.1080/0924345920030404

Rost, D. H., & Hanses, P. (1994). Possessions and use of toys in elementary-school boys and girls: Does giftedness make a difference? *Educational Psychology, 14,* 181–194. doi: 10.1080/0144341940140203

Rumberger, R. W., & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, *107*, 1999–2024.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*, 454–499. doi:10.3102/0034654307310317

Shah, B., Dwyer, C., & Modood, T. (2010). Explaining educational achievement and career aspirations among young British Pakistanis: Mobilizing 'ethnic capital'? Sociology, 44, 1109–1127. doi:10.1177/0038038510381606

Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft [Scholastic achievement of adolescents with migration background: The role of class composition]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* [Origin-related disparities in education: Differential educational processes and the difficulties of distributive justice ] (pp. 189–219). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund [Pupils with a migration background]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Eds.), *PISA 2009: Bilanz nach einem Jahrzehnt* [PISA 2009: Balance after a decade] (pp. 200–230). Münster, Germany: Waxmann.

Statistik Austria. (2013). Registerbasierte Statistiken, Kalenderjahr 2013 [Register-based statistics, year 2013]. Retrieved from https://www.statistik.at/.../registerbasierte_statistiken_2013_-_einkommen_rs_sb_10.6_071115.pdf

Statistik Austria. (2014). Schülerinnen und Schüler mit nicht-deutscher Umgangssprache im Schuljahr 2013/14 [Pupils with non-German mother tongue, year 2013/14]. Retrieved from http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bildung_und_kultur/formales_bildungswesen/schulen_schulbesuch/index.html

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest,12,* 3–54. doi: 10.1177/1529100611418056.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning for early careers in science. *Science, 312,* 1143–1144. doi:10.1126/science

Thapa, A., Cohen, J., Guffey, S., & Higgins-D´Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83,* 357–385. doi: 10.3102/0034654313483907

Thoren, K., Heinig, E., & Brunner, M. (2016). Relative age effects in mathematics and reading: Investigating the generalizability across students, time and classes. *Frontiers in Psychology, 7*, 679. doi:10.3389/fpsyg.2016.00679

Tiedemann, J., & Billmann-Mahecha, E. (2004). Kontextfaktoren der Schulleistung im Grundschulalter. Ergebnisse der Hannoverschen Grundschulstudie [Context factors of primary school performance: Results from the Hanover primary school study]. *Zeitschrift für Pädagogische Psychologie*, *18*, 113–124. doi:10.1024/1010-0652.18.2.113

Van der Slik, F. W. P., Driessen, G., & De Bot, K. (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudinal multilevel examination in Dutch elementary schools. *European Sociological Review, 22,* 293–308. doi:10.1093/esr/jci058

Verhaeghe, J. P., Vanlaar,G., Knipprath, H., De Fraine, B., & Van Damme, J. (2018). Can group composition effects explain socioeconomic an ethnic achievement gaps in primary education? *Studies in Educational Evaluation, 57,* 6–15. doi:10.1016/j.stueduc.2017.07.006

Weiß, R. (2006). *CFT 20-R*. *Manual*. Göttingen, Germany: Hogrefe.

White, S.L.J., Graham, L.J., & Blaas, S. (2018). Why do we know so little about the factors associated with gifted underachievement? *Educational Research Review, 24,* 55–66. doi:10.1016/j.edurev.2018.03.001

Wolak, M. (2016). *Facilitating estimation of the Intraclass Correlation Coefficient.* Retrieved from https://cran.r-project.org/web/packages/ICC/ICC.pdf

Worrell, F. C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly, 53,* 242–244. doi:10.1177/0016986209346828.

Wroblewski, A. (2012). Situation und Kompetenzen von Schüler/inne/n mit Migrationshintergrund: Deskriptive und multivariate Analyse der Determinanten der Testleistung [Situation and competencies of students with a migration background: Descriptive and multivariate analysis of the determinants of test performance]. In F. Eder (Ed.), *Pisa 2009. Nationale Zusatzanalysen für Österreich* [Pisa 2009: Additional national analyses for Austria] (pp. 337–366). Münster, Germany: Waxmann.

Yung, Y.-F., & Chan, W. (1999). Statistical analysis using bootstrapping: Concepts and implementations. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 82–105). Thousand Oaks, CA: Sage Publishing House.

Zhou, M. (2005). Ethnicity and social capital: community-based institutions and embedded networks of social relations. In G. C. Loury, T. Modood, & S. M. Teles (Eds.), *Ethnicity, social mobility and public policy* (pp.131–159). Cambridge, UK: Cambridge University Press.

Ziegler, A., & Stöger, H. (2009). Begabungsförderung aus einer systemischen Perspektive [Fostering giftedness from a systemic perspective]. *Journal für Begabtenförderung*, *9*, 6–31.

Ziegler, A., & Heller, K. (2000). Approach and avoidance motivation as predictors of achievement behavior in physics instructions among mildly and highly gifted eight-grade students. *Journal for the Education of the Gifted, 23,* 343–359. doi:10.1177/016235320002300402

Zimmer, K., Burba, D., & Rost, J. (2004). Kompetenzen von Jungen und Mädchen [Boys´ and girls´ competencies]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, . . . U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* [PISA 2003. The educational background of adolescents in Germany- findings of the second international comparative study] (pp. 211–223). Münster, Germany: Waxmann.

Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management, 19,* 75–92. doi: 10.1002/(SICI)1520-6688(200024)19:1<75::AID-PAM5>3.0.CO;2-W

**Appendix**

**Bootstrapping for Cohort 1**

| | R | Original | BootBias | BootSE | BootMed |
|---|---|---|---|---|---|
| (Intercept) | 10,000 | 54.57 | 0.08 | 7.91 | 56.12 |
| Reasoning | 10,000 | 4.26 | 0.01 | 2.92 | 4.36 |
| Age | 10,000 | 1.36 | 0.01 | 1.03 | 1.34 |
| Gender | 10,000 | 5.78 | 0.00 | 1.55 | 5.91 |
| Self-concept | 10,000 | 3.64 | 0.00 | 0.98 | 3.54 |
| Parental involvement | 10,000 | -5.72 | -0.03 | 1.25 | -5.78 |
| Subject specific interest | 10,000 | -0.88 | -0.01 | 0.84 | -0.91 |
| Immigration status | 10,000 | -0.91 | 0.00 | 1.73 | -0.99 |
| Homogeneity in reasoning skills | 10,000 | -4.25 | 0.06 | 2.81 | -4.26 |
| Heterogeneity in reasonong skills | 10,000 | -5.84 | -0.00 | 1.79 | -5.87 |
| Class atmosphere | 10,000 | 2.25 | -0.00 | 1.29 | 2.25 |
| Immigration ratio | 10,000 | -7.88 | 0.06 | 3.13 | -7.92 |
| Individualization | 10,000 | 0.77 | -0.00 | 1.40 | 0.80 |
| Class size | 10,000 | -7.03 | -0.01 | 1.60 | -7.04 |
| Composition of parental involvement | 10,000 | -2.49 | 0.02 | 2.34 | -2.50 |
| Gender composition | 10,000 | -3.22 | -0.18 | 3.92 | -3.38 |

**Confidence Intervals for Cohort 1**

| | 2.5 % | 97.5 % |
|---|---|---|
| Sig01 | 3.59 | 18.56 |
| Sig02 | -1.00 | -1.00 |
| Sig03 | 2.74 | 12.84 |
| Sigma | 3.80 | 5.53 |
| (Intercept) | 60.55 | 91.52 |

| | | |
|---|---|---|
| Reasoning | -0.99 | 10.60 |
| Age | -0.52 | 3.60 |
| Gender | 2.56 | 8.69 |
| Self-concept | 1.096 | 4.95 |
| Parental involvement | -1.05 | 1.67 |
| Subject specific interest | -8.11 | -3.08 |
| Immigration status | -1.99 | 1.68 |
| Homogeneity in reasoning skills | -9.54 | 1.49 |
| Heterogeneity in reasoning skills | -9.49 | -2.51 |
| Class atmosphere | -0.48 | 4.58 |
| Immigration ratio | -14.06 | -1.88 |
| Individualization | -1.91 | 3.50 |
| Class size | -10.34 | -3.93 |
| Composition of parental involvement | -7.15 | 2.09 |
| Gender composition | -5.46 | -1.02 |

**Bootstrapping for Cohort 2**

| | R | Original | BootBias | BootSE | BootMed |
|---|---|---|---|---|---|
| (Intercept) | 10000 | 48.54 | -0.00 | 1.69 | 49.00 |
| Reasoning | 10000 | 4.72 | -0.01 | 0.95 | 4.72 |
| Age | 10000 | -0.68 | -0.01 | 0.60 | -0.67 |
| Gender | 10000 | 3.78 | 0.00 | 1.02 | 3.76 |
| Self-concept | 10000 | 3.13 | 0.01 | 0.56 | 3.14 |
| Parental involvement | 10000 | -0.65 | 0.00 | 0.54 | -0.64 |
| Subject specific interest | 10000 | 1.30 | -0.01 | 0.52 | 1.29 |
| Immigration status | 10000 | -0.47 | -0.00 | 1.01 | -0.46 |

**Confidence Intervals for Cohort 2**

| | **2.5 %** | **97.5 %** |
|---|---|---|
| Sig01 | 0.30 | 2.62 |
| Sigma | 6.72 | 8.05 |
| (Intercept) | 45.69 | 52.29 |
| Reasoning | 2.85 | 6.63 |
| Age | -1.86 | 0.50 |
| Gender | 1.77 | 5.77 |
| Self-concept | 2.02 | 4.22 |
| Parental involvement | -1.69 | 0.39 |
| Subject specific interest | 0.28 | 2.32 |
| Immigration status | -2.41 | 1.52 |

[1] Data collection in Viennese schools requires school district authorization district and does not permit a total testing time of more than 60 min for students. Therefore, reasoning (sequence completion and matrices sub-scales) needed to be selected from among three factors. Reasoning explained the highest amount of variance (50%).
[2] The CFT 20-R (Culture Fair Test) measures fluid ability, in line with Cattell, using built-in time limits. This non-verbal test can be used with children from 8.5 to 19 years old as well as adults from 20 to 60 years old.
[3] There are no standardized assessments that exactly match the Austrian curriculum.
[4] According to Jones (1992), the residual variance between groups can increase when a Level 1 predictor — here it is the variance in the slope of the reasoning predictor — is added to the model. Of course, the proportion of explained variance for the overall model must then increase as well.
[5] The study was supported by a Talent Austria grant from OeAD GmBH, financed by the Austrian Federal Ministry of Education, Science and Research (BMBWF).