

---

KATJA SCHARENBERG

**Do secondary students learn more in homogeneous or heterogeneous classes?  
The importance of classroom composition for the development of  
Reading achievement in secondary school.<sup>1</sup>**

---

Dr. Katja SCHARENBERG

Institute for School Development Research (IFS)  
TU Dortmund University  
Vogelpothsweg 78  
D – 44227 Dortmund

Germany

[scharenberg@ifs.tu-dortmund.de](mailto:scharenberg@ifs.tu-dortmund.de)

---

<sup>1</sup> This article is based on some of the results of my PhD thesis (Scharenberg, 2012).

**Abstract**

Ability grouping is one of the most controversial issues in educational research (Slavin, 1987). In theory, proponents of ability grouping emphasise that it provides instruction that can specifically be adapted to the students' needs for learning. Opponents of ability grouping, however, argue that ability grouping is especially disadvantageous for low-ability students.

The present study takes up this discussion about the pros and cons of ability grouping and examines whether a homogeneous or heterogeneous classroom composition is more suitable for the development of the students' Reading achievement in grades 7 and 8 in secondary school. Based on previous analyses for grades 5 and 6, it is hypothesised that – controlling for individual variables as well as classroom compositional and institutional variables – student learning is more effective in heterogeneous classes.

The paper presents findings within the context of the longitudinal study KESS in Hamburg, Germany. A total of 5,941 students in 331 classes were included in the analysis sample. Multilevel models show that the strongest predictor of the students' Reading achievement at the end of grade 8 is their individual prior knowledge at the previous point of measurement. On the class level, a heterogeneous classroom composition is neither associated with a higher nor a lower development of achievement. High-ability and high-SES classes, however, prove to be effective for the development of achievement. Commonality analyses show that these compositional characteristics are highly confounded with the different tracks, indicating that the effectiveness of heterogeneous or homogeneous classes also depends on their average achievement and social composition as well as the school track.

**Keywords:** Heterogeneity, ability grouping, classroom composition, multilevel modelling, longitudinal study

### 1. Homogeneous versus heterogeneous classes

Grouping students by their ability is one of the most controversial issues in educational research and politics. There are many theoretical arguments and empirical answers to the question whether learning can be more effective in terms of learning outcomes and students' development of achievement in either homogeneous or heterogeneous classes.

According to Slavin (1987; 1990), the advantages of ability grouping are manifold:

1. Ability grouping permits students to progress according to their specific abilities.
2. Ability grouping is supposed to reduce the students' risk of failures.
3. From the teachers' point of view, it is easier to adapt instruction techniques to the needs and ability level of the group.
4. It makes individual instruction possible so that instruction for high-ability students can take place on a higher level and with an accelerated pace. Low-ability students, however, are supposed to get more individual attention by their teachers and enough time for repetition of the learning content in homogeneous classes.
5. Moreover, ability grouping helps to maintain interest and is an incentive for the high-ability students because they are not bored when they have to learn together with low-ability students. Low-ability students, in contrast, participate more when they are not confronted with too excessive demands by their teachers.
6. Finally, it is assumed that ability grouping makes teaching easier as teachers can gear their instruction to more or less one performance level for all students and do not have to vary in their teaching style and learning tasks as much as in heterogeneous classes.

Opponents of ability grouping, however, focus their contra-arguments primarily on low-ability students who are especially disadvantaged in homogeneous classes. Again referring to Slavin (1987; 1990), the disadvantages of ability grouping are as follows:

1. Low-ability students are put together with equally able students to classes with a lower ability level.
2. However, low-ability students need the presence of high-ability students to be stimulated and encouraged for learning. But in contrast, by ability grouping, low-ability students lack a

positive reference group and are deprived of the stimulation by high-achievers.

3. In homogeneous classes, especially low-ability students often experience a lower quality of instruction and have teachers who are less experienced or less qualified.
4. Moreover, low-ability students are stigmatised which leads to a further discouragement of these students when teachers communicate only low expectations.
5. Finally, teachers object to low-ability classes and prefer teaching in high-ability classes.

To summarise these theoretical pros and cons of ability grouping, it seems that proponents of ability grouping emphasise the higher effectiveness in homogeneous learning groups and thus concentrate more on technical aspects of teaching and learning (Gamoran, 1986). Opponents of ability grouping, however, focus their arguments on effectiveness and efficiency as well as on equity and discrimination of certain student subpopulations (Hattie, 2002; Slavin, 1990).

### 2. A short overview on recent research

Findings from a broad international research on ability grouping, however, show that ability grouping is not only theoretically controversial, but also contentious in empirical educational research and its statistical and methodological approaches.<sup>2</sup> In summary, it can be stated that findings on the effects of ability grouping are inconsistent: Some studies show that the effects of grouping students into homogeneous learning groups are positive (e.g. Dar & Resh, 1986; Luyten & van der Hoeven-van Doornum, 1995; Opdenakker & van Damme, 2001; Sörensen & Hallinan, 1986), other studies conclude that it rather tends to result in a lower development of achievement (e.g. Hoffer, 1992). However, especially meta-analyses find in general no effect of ability grouping at all or only small effects (e.g. Kulik & Kulik, 1982, 1984). Moreover, there seem to be no general, but rather domain-specific effects depending on the type of the subject (e.g. Slavin, 1987, 1990). There is, instead, evidence of differential effectiveness of ability grouping (Kulik & Kulik, 1982, 1984; Slavin, 1987, 1990): Disadvantages of homogeneous ability grouping can be

<sup>2</sup> This is just a very short overview on the main differences between the studies. For a more detailed and systematic overview on studies relating to ability grouping since the 1970s, refer to Baumert et al. (2006), Bellin (2009), Scharenberg (2012) or Schofield (2006).

shown for low-SES students, for students with migration background or low-ability students. Comparing different studies is often difficult because they differ in their operationalization of ability grouping. Most of the studies focus on the average achievement as a compositional characteristic of classes (e.g. in summary Hattie, 2002; Thrupp, Lauder & Robinson, 2002). Only a few studies have yet analysed the effect of the variation or dispersion of achievement within classes (e.g. Reynolds & Teddlie, 2000; Scheerens, 2008). As analysis methods, most of the studies use analyses of variance or linear regression analyses (e.g. Helmke, 1988; Hoffer, 1992; Resh & Dar, 1992). The more recent studies mostly apply hierarchical linear regression analyses accounting for the nested data structure in the school context (e.g. Hanushek, Kain, Markman & Rivkin, 2003; Opdenakker & van Damme, 2001). Finally, the designs of the studies are either cross-sectional or longitudinal (e.g. Hanushek et al., 2003; Kerckhoff, 1986; Kim, Lee & Lee, 2008).

### 3. Creating homogeneous classes in Germany

The handling of the students' diverse abilities differs widely within the European context. In most European countries, students predominantly learn in heterogeneous classes and are only regrouped for selected subjects. In Germany, however, the homogenisation of learning groups can still be considered as the predominant strategy of the educational system (Tillmann & Wischer, 2006). According to Tillmann (2007), typical processes of creating learning groups that are relatively homogeneous with regard to achievement are:

- school enrolment of children of the same age (usually with the completion of the 6<sup>th</sup> year of age),
- postponement of school enrolment for children who do not yet comply with the requirements for school,
- repetition of a whole school year for students who fail,
- tracking at the age of 10 or 12,
- sending students to lower tracks when they do not come up with the teachers' expectations or have already repeated a school year twice.

### 4. Research questions and hypotheses

The analyses presented in this paper relate to the pros and cons of ability grouping within classes in

German secondary schools. They focus on three key questions:

1. Does achievement heterogeneity within classes affect student achievement at all?
2. Do students learn more in homogeneous or heterogeneous classes?
3. How much of the variance in achievement only accounts for achievement heterogeneity and to what extent is it confounded with other compositional and institutional characteristics?

Based on the controversial theoretical discussion on the effects of ability grouping and mostly negative empirical evidence of homogeneous ability grouping, it is assumed that besides individual learning conditions, achievement heterogeneity is a compositional parameter on the class level. Thus for the multilevel analyses, it is expected to find a significant effect of achievement heterogeneity on the class level. According to mostly negative evidence of homogeneous ability grouping – especially for the low-achievers – it is assumed that achievement heterogeneity, in contrast, positively affects individual achievement. Previous studies on the importance of compositional and institutional characteristics on students' achievement (e.g. Baumert et al., 2006; Gröhlich, Guill, Scharenberg & Bos, 2009) showed that – despite their independent specific effects – they both seem to be highly correlated. Thus, it is expected that achievement heterogeneity as a single compositional variable only slightly accounts for the variance in achievement on the student level. Instead, it is assumed that it is highly confounded with other compositional and institutional characteristics such as the mean achievement and social composition or the different school tracks.

### 5. Methodology

#### *Analysing hierarchical data*

The analysis of compositional effects requires a multilevel modelling. Hierarchical linear regression analyses using HLM 6.08 (Raudenbush & Bryk, 2002) are applied to analyse the effects of variables on the student and class level. The dependent variable is the students' Reading achievement at the end of grade 8 (z-standardised on the student level). Predictors on the student level are gender (reference: male), migration background (reference: both parents born in Germany), social background (HISEI; Ganzeboom, De Graff & Treiman, 1992),

general intelligence (cognitive abilities test with non-verbal figural analogies; Heller & Perleth, 2000) and prior knowledge (achievement score in Reading at the beginning of grade 7). Predictors on the class level are achievement heterogeneity (operationalised as the standard deviation of Reading achievement at the beginning of grade 7 aggregated by the classroom affiliation in grade 8), the mean Reading achievement and social composition (both aggregated by the classroom average) as well as the school track (reference: basic track).

For the interpretation of the meaningfulness of the results, three different indicators are applied: *First*, standardised regression coefficients ( $\beta$ ) are reported. In order to estimate the compositional effects directly, all continuous predictors are z-standardised on the student level ( $M = 0$ ;  $SD = 1$ ) and then aggregated on the class level. By doing so, the standardised regression coefficients represent the change in terms of standard deviations in the dependent variable (here: Reading achievement at the end of grade 8) resulting from a change in an independent variable by one standard deviation or by choosing the appropriate reference category of dichotomous variables (e.g. the difference in Reading achievement between boys and girls). *Second*, for an easy interpretation of the results, effect sizes that are comparable to Cohen's  $d$  are reported. According to Tymms (2004), the effect size  $\Delta$  for continuous level-2 predictors can be calculated by twice the product of the unstandardised regression coefficient ( $B$ ) and the ratio of the standard deviation of the predictor variable on the class level ( $SD_{predictor}$ ) and the residual standard deviation on the student level ( $\sigma_e$ ):

$$\Delta = 2 \cdot B \cdot SD_{predictor} / \sigma_e$$

According to Bortz & Döring (1995), effect sizes of  $\Delta = 0.20$  are considered as small effects, whereas effect sizes of  $\Delta = 0.50$  or  $\Delta = 0.80$  constitute medium or large effects. *Third*, the model fit can also be estimated by the proportion of variance ( $R^2$ ) on the student and class level that can be explained by the predictor variables at each of the levels (Snijders & Bosker, 1999). In order to compare the model fits of differently specified models, deviance statistics (e.g. Langer, 2008) are reported.

Commonality analysis (Cohen, West, Cohen & Aiken, 2003) are conducted in order to determine the percentage of variance in achievement that can be accounted for a single predictor variable. By

applying this method, it is possible to partition the regression effects of the level-2 predictors from the multilevel analyses into specific and common effects. Specific effects indicate how much of the variance can only be explained by a unique variable. Common effects, in contrast, indicate the amount of variance in the dependent variable being explained by a set of several variables. With regard to the research questions being examined here, the commonality analysis provides the information how much of the variance in Reading achievement at the end of grade 8 can only be explained by the achievement heterogeneity within the classes and to what extent it is confounded with other compositional and institutional variables.

#### *Missing data*

Missing data are always a challenge in multilevel analyses. For the present dataset, the average percentage of missing data ranges from 1.5 % in the cognitive abilities test and 4.7 % in the achievement scores. The background variables were derived from the student and parent questionnaires with a lower return rate and percentages of missing values ranging between 0.2 % and 53.7 %. As the present study is longitudinal and repeated measurements on the same individuals often tend to be correlated (Schafer & Graham, 2002), missing information could partially be recovered from previous or later measurements. By doing so, the amount of missing data could be reduced to 20.3 % at most in a first step. In a second step, missing data were estimated with the software NORM 2.03 (Schafer, 1999) based on an imputation model using background variables that correlate with achievement ( $|r| \geq .30$ ).

#### **6. Dataset and description of the sample**

Data source for the analyses is the longitudinal study KESS that examines the competencies and attitudes of students in the federal state of Hamburg in Germany. Measurements of achievement were taken at the end of grade 4 (KESS 4), at the beginning of grade 7 (KESS 7) and at the end of grade 8 (KESS 8). Data for the first point of measurement were taken at the end of grade 4 immediately before the transition from primary to secondary school so that the achievement data can also be interpreted as the initial learning situation at the beginning of grade 5. As the second measurement took place immediately after the beginning of the term in

grade 7, achievement data can also be interpreted as the achievement at the end of grade 6. At each point of measurement, an entire cohort of around 14,000 students each has been tested. The longitudinal sample comprises 9,628 students who participated in all the three points of measurement.

Following up previous analyses for the development of achievement from grade 5 to 6 (Bos & Scharenberg, 2010; Gröhlich, Scharenberg & Bos, 2009), the results presented here in this paper focus on the students' further development of achievement in grades 7 and 8. In order to obtain a realistic and valid estimate of classroom compositional variables, only those classes were selected for the analyses with more than 10 students. Furthermore students from comprehensive schools had to be excluded from the analyses as they are divided into different high- and low-ability classes for each subject from grade 7 on. Thus learning groups in comprehensive schools are differently composed and rearranged for each of the main subjects. This leads to a reduction of the analysis sample to 5,941 students in 331 classes, of which 51.9 % are girls and 37.8 % are students with at least one parent not born in Germany. The analysis sample consists of 49 classes at lower tracks (*Hauptschule*), 69 classes at intermediate tracks (*Realschule*) and 213 classes at higher tracks (*Gymnasium*).

Descriptive analyses show that the learning groups at the three tracks differ clearly with regard to their average classroom composition ( $p < .05$  for following differences between the classes): The mean Reading achievement at the beginning of grade 7 in classes at lower tracks is more than one standard deviation ( $-1.23 SD$ ) below, the mean achievement in intermediate tracks about half a standard deviation ( $-0.50 SD$ ) below the average of the whole analysis sample. At higher tracks, in contrast, the mean achievement is about a third of a standard deviation ( $0.36 SD$ ) above the grand mean. Significant differences can also be observed for the average social composition: Classes at higher tracks have the most favourable social composition ( $0.27 SD$  above the average), whereas classes at intermediate ( $-0.49 SD$ ) and lower tracks ( $-0.78 SD$ ) have a more unfavourable social composition compared to the whole analysis sample.

Moreover, classes at the different tracks also significantly differ with regard to their heterogeneity: Classes at higher tracks show – on average – a

higher variation of Reading achievement ( $SD = 0.79$ ) than classes at intermediate tracks ( $SD = 0.73$ ). The difference is, however, small ( $d = 0.36$ ). Classes at higher and lower tracks ( $SD = 0.77$ ) are equally heterogeneous.

## 7. Results

### *The classroom as level of analysis*

At first, the distribution of the variance in the dependent variable is analysed. Thus, it is asked how much of the variance in Reading achievement at the end of grade 8 can be explained by differences between the students and by differences between the classes. Model 0 in table 1 shows that although most of the variance in achievement is situated on the student level, there are substantial differences between the classes: 44.6 % of the variance in individual achievement is attributable to differences between the learning groups. The remaining 55.4 % of the variance account for differences between the students.

For a realistic estimate of compositional effects, however, it is important that the model specification on the student level is correct. In longitudinal studies like KESS, it is especially important to take into account the students' prior knowledge at the previous point of measurement. Controlling for student variables (gender, migration background, HISEI, general intelligence and prior knowledge), the intraclass correlation (ICC) decreases, but remains substantial with 23.4 % of the variance in Reading achievement still being attributable to differences between the students. Thus, the main part of the variance that was originally found between the classes can actually be explained by differences between the students. In the following multilevel analyses, the explained variance on the class level relates to this residual variance. All in all, however, the classroom proves to be the relevant level of analysis.

### *Classroom composition and development of achievement*

For the prediction of the students' Reading achievement at the end of grade 8, the multilevel models are equally specified on the individual level and differently specified on the class level. Classroom variables are sequentially introduced: Model 1 only accounts for the achievement heterogeneity of the classes. Model 2 additionally considers the

mean prior knowledge and the mean HISEI as further indicators of classroom composition. Model 3, instead, considers the school track (reference: basic track) as an institutional variable. Model 4 is fully specified and examines the joint effects of compositional as well as institutional variables.

On the student level, all models show that the most important predictor of Reading achievement at the end of grade 8 is the students' prior knowledge at the beginning of grade 7 ( $0.38 \leq \beta \leq 0.43$ ). Girls achieve significantly higher in Reading than boys ( $0.25 \leq \beta \leq 0.27$ ). Furthermore, a higher general intelligence is correlated with a statistically significant higher Reading achievement ( $0.12 \leq \beta \leq 0.14$ ). Students with at least one parent not born in Germany achieve significantly lower ( $-0.08 \leq \beta \leq -0.06$ ) than students without migration background. On the student level, the predictor variables explain around a third of the individual variance in Reading achievement ( $0.296 \leq R^2 \leq 0.298$ ).

On the class level, model 1 shows that the standard deviation of prior knowledge within classes has no statistically significant effect on student achievement. Thus, no variance can be explained on the class level.<sup>3</sup> Furthermore, as can be seen from the deviance statistics, the consideration of achievement heterogeneity as the only variable on the class level leads to no significant improvement of the model fit compared to model 0 that only controls for individual variables.

Model 2 shows that the effect of achievement heterogeneity is again insignificant when controlling for further compositional variables. Students in high-ability classes reach a significant higher Reading achievement ( $\beta = 0.34$ ;  $\Delta = 0.99$ ) than students in low-ability classes. Classes with a higher-than-average social composition offer a more favourable learning environment than socially less privileged classes ( $\beta = 0.12$ ;  $\Delta = 0.29$ ). Controlling for all compositional variables, the percentage of explained variance on the class level is increased to 32.3 %. The consideration of the mean achievement and average social composition of the classes leads to a significant improvement of the model fit compared to model 1.

In model 3, besides the achievement heterogeneity, the school track as an institutional variable is considered. Again, the effect of achievement heterogeneity is insignificant. Students in intermediate tracks significantly perform higher ( $\beta = 0.40$ ) than students in basic tracks. The advance for students at higher tracks is even higher ( $\beta = 0.79$ ).<sup>4</sup> When controlling for achievement heterogeneity and the school tracks, the percentage of explained variance on the class level increases to 37.9 %. The consideration of the school tracks again leads to a significant improvement of the model fit compared to model 1.

Finally, compositional and institutional variables are jointly considered in model 4. As in the previous models, the standard deviation of prior knowledge within classes has no significant effect on the students' Reading achievement at the end of grade 8. Thus, the students' Reading achievement in heterogeneous classes does not significantly differ from the attainment in homogeneous classes. Compared to model 2, the effect of the average achievement composition of the classes decreases and becomes insignificant when controlling for the school track. The effect of the mean social composition only slightly decreases and is still significant, indicating that students in high-SES classes achieve significantly higher ( $\beta = 0.10$ ) than students in socially less privileged learning environments. But the effect is small ( $\Delta = 0.25$ ). The strongest differences in Reading achievement, however, can be observed for the school tracks: Students at intermediate ( $\beta = 0.31$ ) or academic tracks ( $\beta = 0.55$ ) achieve significantly higher in Reading at the end of grade 8 than students at basic tracks. With  $R^2 = 0.393$ , the percentage of explained variance on the class level is somewhat higher in the fully specified model 4 than in model 3. As can be seen from the mean deviance, however, the model fit is only slightly, but not statistically significant improved when considering all variables on the class level in model 4 compared to model 3.<sup>5</sup>

<sup>3</sup> As the negative proportion of explained variance on the class level is only very small, it can be neglected here. However, this is a well-known phenomenon in methodological research (e.g. Snijders & Bosker, 1994).

<sup>4</sup> Moreover, as can be estimated from the standard errors and confidence intervals, the difference in Reading achievement between classes at intermediate and higher tracks is also statistically significant.

<sup>5</sup> In contrast to model 3, however, the difference in Reading achievement between classes at intermediate and higher tracks is no longer significant when controlling for all compositional variables.

Table 1.

Predicting Reading achievement at the end of grade 8. Results from multilevel modelling

	Model 0	Model 1	Model 2	Model 3	Model 4
Intercept		-0.11 (0.08)	<b>-0.19</b> (0.07)	<b>-0.72</b> (0.08)	<b>-0.57</b> (0.09)
<i>Student level</i>					
Gender <sup>1</sup>		<b>0.27</b> (0.02)	<b>0.26</b> (0.02)	<b>0.25</b> (0.02)	<b>0.25</b> (0.02)
Migration background <sup>2</sup>		<b>-0.06</b> (0.02)	<b>-0.07</b> (0.02)	<b>-0.08</b> (0.02)	<b>-0.07</b> (0.02)
General intelligence		<b>0.14</b> (0.01)	<b>0.13</b> (0.01)	<b>0.12</b> (0.01)	<b>0.12</b> (0.01)
HISEI		<b>0.07</b> (0.01)	<b>0.04</b> (0.01)	<b>0.05</b> (0.01)	<b>0.04</b> (0.01)
Prior knowledge		<b>0.43</b> (0.01)	<b>0.38</b> (0.01)	<b>0.39</b> (0.01)	<b>0.38</b> (0.01)
<i>Class level</i>					
Prior knowledge (SD)		0.01 (0.10)	0.11 (0.09)	0.01 (0.08)	0.04 (0.08)
Prior knowledge (M)			<b>0.34</b> (0.04)		0.09 (0.05)
HISEI (M)			<b>0.12</b> (0.04)		<b>0.10</b> (0.04)
Track <sup>3</sup>					
intermediate				<b>0.40</b> (0.05)	<b>0.31</b> (0.07)
academic				<b>0.79</b> (0.05)	<b>0.55</b> (0.09)
<i>Explained variance</i>					
Student level	0.554	0.296	0.297	0.298	0.298
Class level	0.446 <sup>4</sup>	-0.003	0.323	0.379	0.393
<i>Model fit</i>					
Mean deviance	12,271.53 <sup>5</sup>	12,276.05	12,110.38	12,085.99	12,080.40
Reference model		Model 0 <sup>6</sup>	Model 1	Model 1	Model 3
Modification of model fit <sup>7</sup>		0	+	+	0

References: <sup>1</sup> male, <sup>2</sup> parents born in Germany, <sup>3</sup> basic track.

Continuous predictors on student level z-standardised.

Significant parameters bold ( $p < .05$ ). Standard error in brackets.

<sup>4</sup> When controlling for student variables, the explained variance on the class level is reduced to 23.4 percent.

<sup>5</sup> Mean deviance when controlling for student variables.

<sup>6</sup> Model 0 controlling for student variables.

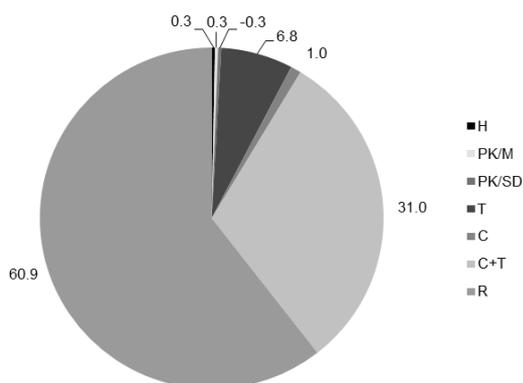
<sup>7</sup> Modifications of model fit compared to reference model insignificant (0) or significant (+).

### Specific and shared variance

Applying a commonality analysis, the relevance of single compositional and institutional variables and their common effects on students' achievement can be illustrated. As figure 1 shows, there is almost no variance in Reading achievement at the end of grade 8 that can be attributed to a single compositional variable.<sup>6</sup> Even taken together as a confounded component, the three compositional variables only explain 1.0 % of the variance in Reading achievement at the end of grade 8. Only the different tracks as indicator of the institutional learning environment explain a detectable proportion of variance (6.8 %). The importance of the compositional variables for the development of Reading achievement in grades 7 and 8 becomes rather noticeable in a confounded component that comprises all of these compositional variables, i.e. the average social composition (mean HISEI), the average achievement composition (mean prior knowledge at the beginning of grade 7) and the heterogeneity of achievement (standard deviation of prior knowledge within the classes), as well as the different tracks as institutional variable. Most of the variance (60.9 %) in Reading achievement at the end of grade 8, however, cannot be explained by the variables that are considered here.

Figure 1.

Frequency of specific and shared variance for the predictors at level 2



H: HISEI (M); PK/M: Prior knowledge (M); PK/SD: Prior knowledge (SD); T: Track; C: Compositional variables (HISEI [M], prior knowledge [M], prior knowledge [SD]); C + T: Compositional variables and track; R: Residual variance.

Frequency in percent.

### 8. Summary and discussion

In contrast to previous analyses within the framework of the KESS study (Hamburg, Germany) that related to the effects of classroom composition for the development of Reading achievement in grades 5 and 6 (Bos & Scharenberg, 2010), the results presented here showed no significant effects of achievement heterogeneity within classes on the development of students' Reading achievement in grades 7 and 8 when controlling for further individual and classroom-related compositional and institutional variables. Thus, the small positive effects of heterogeneous classrooms that were found for the development of Reading achievement from grade 5 to 6 could not be replicated for the students' further development of achievement in secondary school in grades 7 and 8. However, it can be stated that the development of Reading achievement does not significantly differ in homogeneous and heterogeneous classes. This also means that, given the same individual and classroom-related preconditions for learning, a heterogeneous classroom composition does not necessarily seem to be a disadvantage for the development of Reading achievement. The strongest predictors of the development of Reading achievement rather seem to be the students' learning preconditions at the beginning of grade 7 as well as the attended school track.

The commonality analysis showed large proportions of shared variance and almost no proportions of unique variance that is attributable to compositional variables. This means that the effectiveness of heterogeneous classes also depends on the achievement level and the social composition of the classes as well as the school track.

As most of the variance in Reading achievement could not be explained by the variables considered in the analyses here, a desideratum for further analyses would be to include further variables that relate to teaching and instruction to see what can actually make heterogeneous classes effective.

Further analyses within the framework of the KESS study (Scharenberg, 2012) related to – amongst other analyses – differential effectiveness of heterogeneous classes. To answer the question whether high- and low-ability students can equally benefit in heterogeneous classes in grades 5 and 6, slopes-as-outcome models produced a negative interaction effect between achievement heterogeneity on the class level and individual prior knowledge on the

<sup>6</sup> Regarding the negative proportion of explained variance, see p. 7, n. 3.

student level, indicating that the advantages of heterogeneous classes are stronger for the very low-ability students than for high-ability students. High-ability students, however, did not seem to be significantly disadvantaged in heterogeneous classes.

## References

- Baumert, J., Stanat, P. & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bellin, N. (2009). *Klassenkomposition, Migrationshintergrund und Leistung. Mehrebenenanalysen zum Sprach- und Leseverständnis von Grundschulern*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (revised 2<sup>nd</sup> ed.). Berlin: Springer.
- Bos, W. & Scharenberg, K. (2010). Lernentwicklung in leistungshomogenen und -heterogenen Schulklassen. In W. Bos, E. Klieme & O. Köller (eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (pp. 173–194). Münster: Waxmann.
- Cohen, J., West, S. G., Cohen, P. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3<sup>rd</sup> ed.). London: Erlbaum.
- Dar, Y. & Resh, N. (1986). Classroom intellectual composition and academic achievement. A study of the effects of ability-based classes. *American Educational Research Journal*, 23 (3), 357–374.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, 59 (4), 185–198.
- Ganzeboom, H. B. G., De Graff, P. M. & Treiman, D. J. (1992). A standard international socioeconomic index of occupational status. *Social Science Research*, 21 (1), 1–56.
- Gröhlich, C., Guill, K., Scharenberg, K. & Bos, W. (2009). Kumulative Effekte differentieller Lern- und Entwicklungsmilieus von der Jahrgangsstufe 5 bis zur Jahrgangsstufe 8. Paper presented at the 72<sup>nd</sup> Conference of the Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF), Landau.
- Gröhlich, C., Scharenberg, K. & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? *Journal for Educational Research Online*, 1 (1), 86–105.
- Hanushek, E. A., Kain, J. F., Markman, J. M. & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18 (5), 527–544.
- Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37 (5), 449–481.
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+R. Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz.
- Helmke, A. (1988). Leistungssteigerung und Ausgleich von Leistungsunterschieden in Schulklassen: unvereinbare Ziele? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 20 (1), 45–76.
- Hoffer, T. B. (1992). Middle school ability grouping and student achievement in Science and Mathematics. *Educational Evaluation and Policy Analysis*, 14 (3), 205–227.
- Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review*, 51 (6), 842–858.
- Kim, T., Lee, J.-H. & Lee, Y. (2008). Mixing versus sorting in schooling. Evidence from the equalization policy in South Korea. *Economics of Education Review*, 27 (6), 697–711.
- Kulik, C.-L. C. & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19 (5), 415–428.
- Kulik, C.-L. C. & Kulik, J. A. (1984). Effects of ability grouping on elementary school pupils: A meta-analysis. Paper presented at the Annual Meeting of the American Psychological Association. Toronto, Ontario, Canada.
- Langer, W. (2008). *Mehrebenenanalyse. Eine Einführung für Forschung und Praxis* (2<sup>nd</sup> ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- 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 (1), 42–62.
- Opendakker, 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 (4), 407–432.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications. (Advanced quantitative techniques in the social sciences series; vol. 1).
- Resh, N. & Dar, Y. (1992). Learning segregation in junior high-schools in Israel: Causes and consequences. *School Effectiveness and School Improvement*, 3 (4), 272–292.

- Reynolds, D. & Teddlie, C. (2000). The future agenda for school effectiveness research. In C. Teddlie & D. Reynolds (Hrsg.), *The international handbook of school effectiveness research* (pp. 322–343). London: Falmer Press.
- Schafer, J. L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT*. Available: <http://www.stat.psu.edu/~jls/misoftwa.html>.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147–177.
- Scharenberg, K. (2012). *Leistungsheterogenität und Kompetenzentwicklung. Zur Relevanz klassenbezogener Kompositionsmerkmale im Rahmen der KESS-Studie*. Münster: Waxmann. (Empirische Erziehungswissenschaft; vol. 36).
- Scheerens, J. (2008). *Review and meta-analyses of school and teaching effectiveness*. Berlin: Federal Ministry of Education and Research (BMBF).
- Schofield, J. W. (2006). *Migration background, minority-group membership and academic achievement. Research evidence from social, educational, and developmental psychology*. Berlin: Programme on Intercultural Conflicts and Societal Integration (AKI), Social Science Research Center Berlin (WZB). (AKI Research Review 5).
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57 (3), 293–336.
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60 (3), 471–499.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multi-level analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.
- Sørensen, A. B. & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. *American Educational Research Journal*, 23 (4), 519–542.
- Thrupp, M., Lauder, H. & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37 (5), 483–504.
- Tillmann, K.-J. (2007). Lehren und Lernen in heterogenen Schülergruppen. Forschungsstand und Perspektiven. In H. Buchen, L. Horster & H.-G. Rolf (eds.), *Heterogenität und Schulentwicklung* (pp. 7–20). Stuttgart: Raabe.
- Tillmann, K.-J. & Wischer, B. (2006). Heterogenität in der Schule. Forschungsstand und Konsequenzen. *Pädagogik*, 59 (3), 44–48.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (eds.), *But what*

*does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.

Received December 30, 2012