Cognitive Development 44 (2017) 98–109

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Research Paper

What happens if the same curriculum is taught in five instead of six years? A quasi-experimental investigation of the effect of schooling on intelligence

Sebastian Bergold[a,*], Linda Wirthwein[a], Detlef H. Rost[b,c], Ricarda Steinmayr[a]

[a] Technical University Dortmund, Department of Psychology, Emil-Figge-Straße 50, D-44227 Dortmund, Germany
[b] Southwest University Chongqing, Faculty of Psychology, 2 Tiansheng Road, Beibei District, Chongqing 400715, PR China
[c] Philipps-University Marburg, Department of Psychology, Gutenbergstraße 18, D-35032 Marburg, Germany

ARTICLE INFO

Keywords:
Duration of school attendance
Effects of schooling
Intelligence
IQ
Intellectual development
Quasi-experiment

ABSTRACT

Several studies have shown that each year of schooling has a significant positive impact on IQ. However, there is still a debate about how to explain this effect. With the German G8 school reform, the duration of school attendance in the highest track of Germany's tracked school system (Gymnasium) was shortened from 9 years (G9) to 8 (G8) while simultaneously both the curricular contents and amount of instruction should be preserved in full. In the present paper, the G8 reform was utilized as a natural quasi-experiment to examine whether the duration of school attendance would enhance intelligence test scores even when students with different numbers of years of schooling had completed the same curricular contents within the same amount of instruction. Two studies were conducted. In Study A, the performances of $n = 81$ G8 10th graders ($M_{age} = 15.16$ years, $SD = 0.37$) and $n = 80$ G9 11th graders ($M_{age} = 16.39$ years, $SD = 0.49$) on the Berlin Intelligence Structure Test were compared. In Study B, the cognitive abilities of $n = 244$ G8 10th graders ($M_{age} = 15.23$ years, $SD = 0.42$) and $n = 204$ G9 11th graders ($M_{age} = 16.33$ years, $SD = 0.47$) were measured with the Intelligence-Structure-Test 2000 R. The G9 students outperformed the G8 students on nearly all cognitive ability tests despite completing equal curricula. Thus, the impact of schooling on intelligence test scores seems to be primarily due to the fostering of intelligence-related abilities that are independent of formal curricula.

1. Introduction

Schooling has a significant impact on many areas of life, such as labor market participation and income (Schnittker & Behrman, 2012), health (Amin, Behrman, & Spector, 2013; Kemptner, Jürges, & Reinhold, 2011), and the development of social competencies (Bowles & Gintis, 2002). Previous studies have also shown that schooling influences students' intelligence test performance (e.g., Brinch & Galloway, 2012; Cahan & Cohen, 1989; Hansen, Heckman, & Mullen, 2004; Stelzl, Merz, Ehlers, & Remer, 1995). However, the processes underlying this causal relation are still a subject of debate. In the present study, we utilized the German G8 school reform as a natural quasi-experiment to shed some light on these processes. More specifically, we tried to answer the question of whether the impact of schooling on intelligence test scores is due to the teaching of specific cognitive skills and abilities that are tied

* Corresponding author.
E-mail addresses: sebastian.bergold@tu-dortmund.de (S. Bergold), linda.wirthwein@tu-dortmund.de (L. Wirthwein), rost@uni-marburg.de (D.H. Rost), ricarda.steinmayr@tu-dortmund.de (R. Steinmayr).

http://dx.doi.org/10.1016/j.cogdev.2017.08.012
Received 18 December 2015; Received in revised form 6 June 2017; Accepted 25 August 2017
0885-2014/

to the curriculum or to the fostering of more general, intelligence-related abilities that are independent of formal curricula.

1.1. Schooling and intelligence test scores

For quite some time, it has been known that there are substantial correlations between schooling and students' intelligence test scores, mostly ranging from $r = 0.50$ to $r = 0.60$ (Ceci, 1991; Rost, 2013). However, the possibility that schooling influences intelligence has long been seen as rather unlikely. Correlations used to be explained either by selection effects (more intelligent students attend school longer) or by confounding variables that influence both schooling and intellectual development (e.g., socioeconomic status; Ceci, 1991). Meanwhile, studies have suggested that intelligence is not just a precondition but also a consequence of schooling (Brinch & Galloway, 2012; Cahan & Cohen, 1989; Cliffordson & Gustafsson, 2008; Merz, Remer, & Ehlers, 1985; Rost & Wild, 1995; Stelzl et al., 1995).

On the basis of numerous studies, Ceci (1991) suggested that every school year that is missed causes a loss of 0.25–6 IQ points. However, an accurate determination of the effect of schooling on intelligence appears difficult because schooling is confounded with a number of variables that are also relevant for intellectual development, such as physical maturation or intellectual stimulation outside of school (family climate, mass media, etc.), all of which are—just like schooling—indexed by chronological age.

For ethical reasons, it is of course not possible to conduct randomized experiments in which the duration of school attendance is the grouping variable. However, there have been some well-designed quasi-experimental studies that have held chronological age constant in order to isolate the effect of schooling. Cut-off dates for school enrollment allow for an estimation of the schooling effect because students of roughly the same chronological age but with different durations of school attendance can be directly compared on their intelligence test scores (e.g., Merz et al., 1985). Other studies have used the regression discontinuity design (Cahan & Cohen, 1989) in which the intelligence test score is separately regressed on age in each of at least two successive grades. Thus, the slope indicates the age effect, whereas the difference between the predicted values of the oldest students in the initial grade and the youngest students in the subsequent grade represents the schooling effect. Other studies have spread intelligence assessments across an entire school year in order to separate the schooling effect from the age effect as well as from their interaction (Rost & Wild, 1995). Whereas these methodologically sophisticated studies consistently found that schooling has a noteworthy impact on intelligence test scores, they varied with regard to the size of the effect, even within similar age groups and when studying the same intellectual ability. For example, Stelzl et al. (1995) found that one year of schooling accounted for 5.9–8.6 IQ points in fourth graders, depending on the intellectual ability under study. For general intelligence, the effect was 7.6 IQ points (see also Merz et al., 1985). Rost and Wild (1995), also studying general intelligence and drawing on a large sample of third graders, found that every month of schooling accounted for about 0.4 IQ points (which resulted in about 4 IQ points per school year). In a comprehensive review, Rindermann (2011) estimated the average schooling effect on intelligence at 5.6 IQ-points.

However, even though these studies found differing sizes of the effect of schooling, they commonly showed that the effect of schooling was at least twice as large as the effect of chronological age. Some methodologically sound studies even found that chronological age had no noticeable effect on intelligence test scores. In the study by Rost and Wild (1995), the effects of both chronological age and its interaction with duration of school attendance were negligible. Merz et al. (1985) and Stelzl et al. (1995) compared their estimated effects of one year of schooling with the total increase of IQ scores in one year calculated from the norm tables of the intelligence tests they had used. With regard to fluid intelligence, the IQ increase was completely due to schooling. Cliffordson and Gustafsson (2008) examined the magnitudes of both the schooling effect and the age effect on intelligence on the basis of a sample of 48,269 Swedish men. Participants were tested in one of several possible testing sessions for enlisting in military services. Several possible testing sessions were offered across the year in which participants turned 18. The participation date was chosen randomly and was therefore independent of exact age. Thus, both age (as measured by the number of days before and after the 18th birthday, respectively, within this year) and schooling (as measured by the number of days of schooling within this year) varied independently from one another. Whereas the average schooling effect turned out to be 2.7 IQ points, the average age effect was 0 IQ points. These effects remained unchanged even when numerous confounding variables (e.g., socioeconomic status, educational level, ethnicity, or school achievement) were controlled for.

1.2. What underlies the causal link between schooling and intelligence test performance?

As noted above, previous research has convincingly shown that schooling enhances performance on intelligence tests. However, there is still an ongoing debate about why this is the case (i.e., what the processes underpinning this causal relation are). Empirical findings are scarce and partly inconsistent.

On the one hand, there is the hypothesis that the teaching of rather specific abilities that are closely linked to curricula (such as content-specific knowledge or reading ability) is the driving factor behind school causing higher scores on intelligence tests because those abilities can help a person complete an intelligence test successfully (see Ceci, 1991; Neisser et al., 1996; Van de Vijver & Brouwers, 2009). Thus, better scores would be achieved indirectly via those abilities and would therefore not be a consequence of a "real" enhancement of intelligence (although some specific abilities such as reading ability might contribute to intellectual development in the long term; see Ritchie, Bates, & Plomin, 2015). Several studies have shown that schooling indeed fosters specific abilities that are tied to the curriculum, such as reading ability, certain specific mathematical skills, and specific knowledge (Alexander & Martin, 2004; Baltes & Reinert, 1969; Bisanz, Morrison, & Dunn, 1995; Blair, Gamson, Thorne, & Baker, 2005; Crone & Whitehurst, 1999; Cunningham & Carroll, 2011; Naito & Miura, 2001). Cliffordson and Gustafsson (2008) found that, at least in part, there were differential effects of schooling in the academic tracks they investigated, thus suggesting a nontrivial role of the curriculum. More specifically, the authors found an influence of schooling on the performance in a technical comprehension test only in the natural science track and in the technology track. By contrast, this effect could not be demonstrated in either the social science track or the economics track. Bisanz et al. (1995) even found that the effect of schooling on mental arithmetic was limited to the specific tasks that were taught in school.

On the other hand, there is the hypothesis that schooling enhances intelligence test scores because schooling fosters more general, intelligence-related abilities such as abstract thinking and problem solving ability, which are context-free and independent of curricula, that is, not explicitly but incidentally taught in school (see Baker, Salinas, & Eslinger, 2012; Ceci, 1991; Neisser et al., 1996; Scribner & Cole, 1973). Thus, changes in intelligence test scores caused by schooling would then reflect a “real” change in intelligence. Consistent with this explanation, studies have found that schooling indeed fosters more general cognitive abilities such as conditional reasoning (Artman & Cahan, 1993; Artman, Cahan, & Avni-Babad, 2006; Cahan & Artman, 1997; Cahan, Greenbaum, Artman, Deluya, & Gappel-Gilon, 2008) and short-term memory (Morrison, Smith, & Dow-Ehrensberger, 1995).

Studies that have investigated these two hypotheses against each other are scarce and their findings have been ambiguous. Using a hierarchical model of intelligence, Gustafsson (2008) found that schooling had the largest effect on fluid intelligence, suggesting a “real” change in intelligence. By contrast, results from a “natural” quasi-experiment with children in Malawi suggested that there was no strong effect of schooling on basic cognitive processes (Van de Vijver & Brouwers, 2009), suggesting that the impact of schooling on intelligence test scores is rather artificial. Ritchie, Bates, and Deary (2015) investigated whether years of education were associated with changes in the *g* factor or in more specific cognitive abilities. Three models were fitted. The model displaying the best fit was a model with paths from years of education to the specific abilities and without a path from years of education to *g*, suggesting that schooling impacts specific abilities and not *g*. However, in this study, intelligence was first assessed at age 11 and then assessed again not until age 70. As also the authors note, the long time span might have allowed for a variety of environmental or biological factors to dissipate a possible effect of schooling on *g*. Thus, an effect on *g* might still be found at younger ages.

In the present study, we utilized the German G8 school reform (see below) as a natural quasi-experiment in order to directly test the two hypotheses against each other and thereby shed further light on the processes underlying the impact of schooling on intelligence test performance.

1.3. The German high school (Gymnasium) and the G8 school reform

Germany has a multitracked school system. After 4 years of elementary school, students can attend the Hauptschule, the Realschule, the Gymnasium, or two kinds of comprehensive schools that combine all of these school tracks (the Gemeinschaftsschule and the Gesamtschule). The Gymnasium is the highest track in Germany’s secondary school system. It is the option most frequently chosen for receiving the Abitur, which allows for university enrollment. In 2012/2013, for instance, 44.5% of the students in the German secondary school system attended the Gymnasium (Statistisches Bundesamt, 2014).

About 10–15 years ago, most federal states in Germany reformed the length of time students attended Gymnasiums. Whereas before the reform, schooling at Gymnasiums lasted 9 years until the general leaving examination Abitur (G9), it lasted 8 years after the reform (G8). The reasons for the implementation of G8 were primarily economic in nature (see Kühn, van Ackeren, Bellenberg, Reintjes, & im Brahm, 2013). With the shorter length of the program, the average age at which citizens entered the labor force was lower. This was implemented as a reaction to the demographic changes that German society continues to undergo. An additional goal was to provide a better fit with other countries’ educational systems.

Schooling in the Gymnasium involves two phases, and the shortening of school attendance was applied exclusively to the school years in the first phase. The first phase is the secondary level I (Sekundarstufe I; G9: spanning from Grades 5–10, i.e. 6 years; G8: spanning from Grades 5–9, i.e. 5 years), where students are taught in fixed classes. Then, students change over to the secondary level II (Sekundarstufe II), lasting for 3 years both before and after the reform and involving a course system. Thus, with the G8 reform, the change in length involved only the secondary level I. The duration of the secondary level II remained unchanged (G9: spanning from Grades 11–13; G8: spanning from Grades 10–12). Furthermore, the G8 reform was not accompanied by a reform of the curricular contents. As before the G8 reform, the curricular contents were centrally prescribed and mandatory for all Gymnasiums and remained unchanged after the G8 reform. Therefore, G8 students still had to complete the same curriculum as G9 students but in 8 years instead of 9. This point has always been criticized but was never changed (Kühn et al., 2013). To compensate for this mismatch, the number of lessons per week was increased for G8 students in Grades 5–9. Before the reform, lessons were never or only rarely extended into the afternoon, but this extension became common practice in the secondary level I after the reform. In this way, the overall number of lessons taught before the transition to the secondary level II remained unchanged. The same was also true for the teaching materials (e.g., textbooks). Thus, G8 students have 1 year less schooling compared with G9 students before they enter the secondary level II, but they are expected to learn the same lessons in school in both the number of lessons and the contents.

Students’ parents had no choice about whether their children would complete the Gymnasium in 8 or 9 years. In the federal state of North Rhine-Westphalia where we conducted the current study, all students who entered the Gymnasium (5th grade) in 2004 were scheduled to complete the Gymnasium in 9 years, whereas those who started in 2005 were scheduled to complete it in 8 years. In 2010 (i.e., at the transition from the secondary level I to the secondary level II), the two cohorts were combined into one grade (i.e., the first year of the secondary level II), whereas up to that point, they had been taught the same curriculum (in both the number of lessons and the contents) but with different numbers of years of schooling. We investigated the students right after they had been unified. Hence, the G8 school reform gave us the opportunity to investigate the nature of the effect of schooling on intelligence test scores.

Research regarding the differences between G8 and G9 students has been scarce and has primarily focused on several psychosocial variables (e.g., stress level, health, leisure-time behavior; see Milde-Busch et al., 2010; Trautwein, Hübner, Wagner & Kramer, 2015). In sum, no substantial differences between the two groups of students were found. However, we are not aware of any study that focused on comparing the intelligence of the G8 and G9 students.

1.4. The present study

As outlined above, schooling has been shown to impact intelligence test scores (Brinch & Galloway, 2012; Cahan & Cohen, 1989; Merz et al., 1985; Rost & Wild, 1995; Stelzl et al., 1995; see Rindermann, 2011, and Rost, 2013, pp. 444–452, for a comprehensive review). However, less is known about the processes that underlie this causal relation, that is, whether teaching curricular contents or whether fostering more general cognitive abilities are responsible for the effect of schooling on intelligence test performance. Finding an answer to this question is of high relevance because it could help to clarify whether duration of school attendance actually fosters "real" intelligence or whether better scores on intelligence tests can be explained by specific competencies taught at school that are beneficial when completing an intelligence test. There is some evidence for both explanations: Schooling fosters specific abilities that are tied to curricula and that could enhance intelligence test scores (e.g., Alexander & Martin, 2004; Bisanz et al., 1995; Cunningham & Carroll, 2011). At the same time, schooling also fosters abilities that are both related to intelligence and independent of curricula, such as conditional reasoning (Artman & Cahan, 1993; Artman et al., 2006; Cahan & Artman, 1997; Cahan et al., 2008) and short-term memory (Morrison et al., 1995).

The German G8 school reform provides an opportunity to test these two explanations against each other directly. With the G8 reform, the duration of school attendance was shortened by 1 year while simultaneously both the amount of hours students were taught at school and the curricular contents should be preserved. If the impact of schooling on intelligence test scores could be fully accounted for by the teaching of curricular contents or the amount of hours spent at school, no differences in intelligence test scores between the G8 and G9 students should emerge because the two groups completed the same curriculum. However, if a difference in intelligence test scores were to be found between the G8 and G9 students (which should be in favor of the G9 students), this difference could be attributed to the fostering of general cognitive abilities at school.

We conducted two studies (A and B) in which we utilized the G8 reform as a quasi-experimental treatment. In each study, the intelligence test performances of a sample of G8 students and a sample of G9 students were compared after they had completed the same curriculum but had attended school for different lengths of time (i.e., 8 years vs. 9 years, respectively).

2. Study A

2.1. Method

2.1.1. Participants

The initial sample consisted of $N = 175$ students (56% girls) in their first year of the secondary level II from an urban Gymnasium in North Rhine-Westphalia, a federal state in Germany. Testing was conducted in autumn 2010. The sample comprised G8 and G9 students. As already reported above, the G8 reform had been implemented in North Rhine-Westphalia in the school year that began in the fall of 2005. In autumn 2010, the G8 students who had enrolled in the Gymnasium in 2005 and the G9 students who had enrolled in the Gymnasium in 2004 both began the secondary level II of the Gymnasium. Thus, our participants consisted of the last cohort of G9 students and the first cohort of G8 students in this Gymnasium. There was no opportunity for students or their parents to choose between G8 and G9. The G9 students had entered Grade 11 (i.e., the first year of the secondary level II with G9) after finishing Grade 10. At the same time, the G8 students had progressed from Grade 9 to Grade 10, which was now also the first year of the secondary level II after the G8 reform. Because of an increase in the number of lessons per week for G8 students in the secondary level I, in the fall of 2010, both student groups had completed the same curriculum and both had 3 more years of schooling until the Abitur. Therefore, they were taught together (i.e., in the same courses) from the very beginning of the secondary level II as was the common practice in North Rhine-Westphalia when the transition from G9 to G8 took place. We tested the students about 1 month after they had been unified.

To hold the years of schooling constant between the G8 and G9 students, 7 G8 and 7 G9 students were excluded from the analyses because they had repeated (3 G8 and 2 G9 students) or skipped a grade (4 G8 students and 5 G9 students), as this would have resulted in a different duration of schooling for these students. Thus, the final sample consisted of $n = 81$ G8 students ($M_{age} = 15.16$ years, $SD_{age} = 0.37$; 55.6% girls) and $n = 80$ G9 students ($M_{age} = 16.39$ years, $SD_{age} = 0.49$; 55% girls). When the testing took place, the girls-to-boys ratio among students in German Gymnasiums was about 55:45 (Statistisches Bundesamt, 2012). Thus, the proportions of girls and boys in the present samples of G8 and G9 students were representative of these student populations.

Testing took place during a regular school day and was conducted in groups of about 20 students. Trained students and research assistants administered the tests according to standardized instructions. Participation was voluntary, and students were allowed to take part only if their parents had completed written consent forms. The participation rate in both student groups was about 90%. Students who were not present when testing took place were absent due to illness or for other reasons that were not related to our investigation.

2.1.2. Measures

We administered the Berlin Intelligence Structure Test (Berliner Intelligenzstruktur-Test; BIS test), Form 4 (Jäger, Süß, & Beauducel, 1997), a carefully constructed IQ test with good psychometric properties. The BIS test assesses the intellectual abilities proposed by the Berlin Intelligence Structure Model (Jäger, 1984). According to this model, a faceted approach is applied to combine 4 cognitive operations (operation speed, memory, creativity, and processing capacity) with 3 types of cognitive contents (verbal, numerical, and figural). This leads to 12 distinguishable intellectual abilities. In our analysis, we did not inspect differences between the G8 and G9 students in every one of these possible combinations but confined ourselves to inspecting differences in the operations and in the contents at the general level. The composite score (which subsumes all of the operations and contents) indicates general intelligence (*g*). The BIS test is frequently applied in German-speaking countries and is considered to be grounded in theory and supported by empirical research. Its validity has been demonstrated in particular for the application with adolescents from higher school tracks (see, e.g., Beckmann & Guthke, 1999).

All students also reported their sex, age, and their parents' highest educational levels. The following categories of schooling and professional qualifications were identified. Schooling: 0 = no graduation, 1 = lower secondary education (Hauptschulabschluss), 2 = secondary school certificate (Mittlere Reife), 3 = entrance qualification for university of applied sciences (Fachhochschulreife), 4 = Abitur. Professional qualifications: 0 = no qualification, 1 = apprenticeship (Lehre), 2 = master's certificate/technician (Meister/Techniker), 3 = university of applied sciences degree (Fachhochschulabschluss), 4 = university degree (Hochschulabschluss), 5 = doctoral dissertation (Promotion). As an index of parental educational level, we used the overall sum of the highest schooling and the highest professional qualification of both parents.[1] Mothers' and fathers' educational levels were substantially correlated ($r = 0.39$, $p < 0.001$). The possible range of the composite score was 0–18.

Furthermore, the schools provided us with the students' grade point averages (GPAs) from their last report cards (made anonymous) before their transition to the secondary level II. The GPAs were the averages of the individual grades in mathematics, German, foreign languages, social sciences, and science. In Germany, school grades range from 1 (*very good*) to 6 (*very poor*), so that lower values indicate better achievement. It should be noted that school grades as defined by the Standing Conference of Ministers of Education and Cultural Affairs for the Federal States in the Federal Republic of Germany (KMK) (1968) refer to objective criteria, representing the degree to which the teaching and learning objectives have been reached. Therefore, we used them as an indicator of the students' mastery of the curriculum.

2.1.3. Analyses

Because the G8 students were on average about 1 year younger than the G9 students, we used the chronological ages of all students as a covariate in our analyses. A further variable that is important when examining schooling and intelligence is parents' educational level (e.g., Rost, 2013, Chapter 5; Steinmayr, Dinger, & Spinath, 2010). The educational levels of the parents of the G8 and G9 students were not statistically significantly different ($M_{G8} = 10.00$, $M_{G9} = 9.26$), $F(1, 159) = 1.58$, $p = 0.21$, $d = 0.20$. Nevertheless, there was a small effect in favor of the G8 students' parents. Therefore, we tested whether additionally controlling for parental educational level would have an effect on our results. This was not the case. Therefore, we report the results obtained when only chronological age was inserted as a covariate.

In recent years, structural equation modeling has been increasingly applied when investigating group differences in intelligence, that is, to perform multi-group analyses of mean and covariance structures (MG-MACS) or multiple indicators-multiple causes (MIMIC) models (see Study B). For Study A, however, the sample size was too small to perform either of them. Therefore, we extracted the latent abilities by principal component analyses (PCAs). All subtests affiliated to the respective latent ability were then summed up, with their loadings on the first unrotated PC used as weights (see Rost, 1993, 2013). First, we performed a PCA for the operations and one for the contents. After that, we performed a PCA using all operations and contents factor scores to extract *g* (z-standardized factor-scores with $M = 0$ and $SD = 1$).

We then analyzed group differences in the factor-scores with analyses of variance (MANOVAs and ANOVAs) and covariance analyses (MANCOVAs and ANCOVAs). Type of schooling (G8 vs. G9) was the independent variable, and the BIS factor scores were the dependent variables. We first conducted MANOVAs and ANOVAs (i.e., without control for chronological age) and then MANCOVAs and ANCOVAs with chronological age as a covariate to figure out which effect controlling for age had on the results. Because the items measuring cognitive operations and cognitive contents intersected, we separately conducted a MAN(C)OVA for the operations and a MAN(C)OVA for the contents. When the results were statistically significant, we subsequently conducted univariate AN(C)OVAs. Testing for a group difference in *g* was implemented with a separate AN(C)OVA. As effect size, we used Cohen's *d*. Cohen's *d* was derived from the η^2 values from the AN(C)OVA. Values of $d \geq 0.8$ were considered to indicate a large effect, $d \geq 0.5$ a moderate effect, and $d \geq 0.2$ a small effect (Cohen, 1988).

2.2. Results and summary

Before conducting our main analyses, we tested whether there was a difference between the G8 and the G9 students in their mastery of the curriculum. To this end, we inspected their GPAs. There was no statistically significant difference between the G8 and the G9 students ($M_{G8} = 2.78$, $M_{G9} = 2.62$), $F(1, 158) = 2.73$, $p = 0.10$, $d = 0.26$.[2]

[1] In case that only one parent lived together with the student (i.e., in the same household), we asked the students to also indicate the educational level of the parent not living in the same household. The educational levels were then calculated summing up both values. We did so because even if one parent does not live in the same household, he/she should have impact on the students' socioeconomic status, for example via alimony.

[2] One might argue that relying on school grades to inspect the students' mastery of the curriculum might be problematic because the grades—despite the resolution made by the KMK (1968)—might at least in part have been assigned with reference to a social norm (i.e., the class or the grade). Indeed, the G8 and the G9 students had not been taught together in secondary level I so that both groups built up different social norm groups. However, about 3 years after the assessments for the present

Table 1
Study A: Means (M) and Standard Deviations (SD) of the Z-Standardized Factor-Scores from the Berlin Intelligence Structure Test (BIS) Scales for G8 and G9 Students and Results of the AN(C)OVAs (Covariate: Chronological Age).

	G8		G9		ANOVA			ANCOVA		
	M	*SD*	*M*	*SD*	$F_{(1,159)}$	*p*	*d*	$F_{(1,158)}$	*p*	*d*
BIS contents										
–Verbal ability	−0.22	0.85	0.22	1.09	7.65	0.005	0.45	6.37	0.013	0.40
–Numerical ability	−0.21	0.92	0.22	1.03	7.47	0.006	0.44	11.51	0.001	0.54
–Figural ability	−0.28	0.80	0.28	1.11	12.34	< 0.001	0.58	5.80	0.017	0.38
BIS operations										
–Operation speed	−0.30	0.76	0.30	1.12	15.67	< 0.001	0.63	10.37	0.002	0.51
–Memory	−0.21	0.77	0.21	1.16	7.20	0.008	0.42	1.69	0.196	0.21
–Creativity	−0.32	0.98	0.32	1.02	0.17	0.685	0.06	2.64	0.106	0.26
–Processing capacity	−0.23	0.88	0.23	1.06	9.02	0.003	0.48	6.37	0.013	0.40
BIS General intelligence *g*	−0.28	0.77	0.29	1.12	14.02	< 0.001	0.59	10.45	0.001	0.51

Note. The *d* values were calculated from the η^2 values; positive *d* values indicate differences in favor of the G9 students.

The MANOVAs showed a statistically significant effect on the BIS content factor scores, $F(3, 157) = 4.94$, $p = 0.003$, $d = 0.61$, as well as on the BIS operations, $F(4, 156) = 5.59$, $p < 0.001$, $d = 0.76$. Table 1 (Columns 2–8) shows the means and standard deviations of the G8 and G9 students' factor scores as well as the ANOVA results. The G9 students scored higher than the G8 students on all BIS contents ($0.44 \leq d \leq 0.58$). Furthermore, the G9 students had significantly higher factor scores on the BIS operations operation speed ($d = 0.63$), memory ($d = 0.42$), and processing capacity ($d = 0.48$). The G9 students also displayed a higher *g* than the G8 students ($d = 0.59$).

However, differences in the G8 and the G9 students' ages had not been taken into account in that analysis. Now we did so by considering the students' ages in (M)ANCOVAs. The MANCOVAs revealed a statistically significant G8/G9 effect on the BIS contents, $F(3, 156) = 4.18$, $p = 0.007$, $d = 0.57$, as well as on the BIS operations, $F(4, 155) = 3.48$, $p = 0.009$, $d = 0.60$. The age effect was not statistically significant and rather small in size for the BIS contents, $F(3, 156) = 1.75$, $p = 0.16$, $d = 0.36$, the BIS operations, $F(4, 155) = 1.31$, $p = 0.27$, $d = 0.21$, and the BIS total score, $F(1, 158) = 1.59$, $p = 0.19$, $d = 0.21$. The results of the ANCOVAs are summarized in Columns 9–11 of Table 1. As can be seen, most differences were somewhat reduced after controlling for age, but were still substantial. The G9 students still outperformed the G8 students in all BIS contents ($0.38 \leq d \leq 0.54$), in operation speed ($d = 0.51$), and in processing capacity ($d = 0.40$). There were no statistically significant differences in memory and creativity but at least small effect sizes in favor of the G9 students ($d = 0.21$ and $d = 0.26$). The difference in *g* was also reduced, but was still substantial ($d = 0.51$). Thus, about 86% (0.51 of 0.59) of the overall group difference in *g* were due to the 1 year difference in school attendance, whereas about 14% were due to differences in age.

To sum up, despite completing the same curriculum, and despite showing the same level of curricular achievement, the G9 students displayed higher cognitive abilities as measured by the BIS test, and most of the differences were statistically significant. The effect sizes were small to medium after controlling for age. Thus, these results provide support for the hypothesis that improved intelligence test performance due to schooling is the consequence of enhancing intelligence-related abilities. In the light of these findings, it can be tentatively concluded that curriculum completion is not the main driving factor behind the effect of schooling on intelligence test performance. Nevertheless, to replicate the results of Study A and provide more support for our conclusions, we conducted an additional study to investigate another sample of G8 and G9 students in which we administered a different psychometrically sound intelligence test.

3. Study B

3.1. Method

3.1.1. Participants

The initial sample that we used in Study B consisted of 473 students (58.1% girls) who, as in Study A, were all attending the first year of the secondary level II from another urban Gymnasium in North Rhine-Westphalia. The sample of G9 students was comprised of two consecutive cohorts. The first G9 cohort was tested in autumn 2007, the second in autumn 2008. The sample of G8 students

(*footnote continued*)
study, the schools provided us with the G8 and the G9 students' GPAs from their Abitur exams (i.e., after completing secondary level II; Abitur grades from 27 students were missing, but this drop-out was not systematic with regard to the affiliation to the G8 or the G9 group, $\chi^2(1) = 1.04$, $p = 0.31$). In secondary level II, the G8 and G9 students *had* been taught together, so that the reference norm was the same for all students. There was no statistically significant difference in the Abitur grades between the G8 and the G9 students, and the small numerical difference was in the same direction as after secondary level I ($M_{G8} = 2.40$, $M_{G9} = 2.31$), $F(1, 132) = 0.85$, $p = 0.36$, $d = 0.16$. Furthermore, our finding is in line with the study by Trautwein et al. (2015). This study drew on representative samples of G8 and G9 students and found no substantial differences between G8 and G9 students in both their Abitur grades and scholastic achievement as measured by achievement tests. Taken together, these points hint to the validity of the GPAs provided after secondary level I.

was also comprised of two consecutive cohorts. The first G8 cohort was tested in autumn 2012, the second in autumn 2013. Thus (as opposed to Study A), the G8 and G9 students were not combined into one grade.[3] Nevertheless, as in Study A, the G8 students had completed the same curriculum as the G9 students but were given one year less time to complete the curriculum.

Again, testing took place during a regular school day in groups of about 20 students. Participation was voluntary, and students were allowed to take part only if their parents had completed written consent forms. The participation rate in both student groups was about 93%. There was no evidence for systematic drop-outs.

We again excluded students from the analyses if they had repeated or skipped a grade (14 G8 and 4 G9 students had repeated a grade, 6 G8 students and 1 G9 student had skipped a grade). Thus, the final sample consisted of $n = 244$ G8 students ($M_{\text{age}} = 15.23$ years, $SD_{\text{age}} = 0.42$; 61.5% girls) and $n = 204$ G9 students ($M_{\text{age}} = 16.33$ years, $SD_{\text{age}} = 0.47$; 55.4% girls). The proportions of boys and girls were roughly representative (Statistisches Bundesamt, 2009, 2010, 2014, 2015).

3.1.2. Measures

We administered the basic module of the Intelligence-Structure-Test 2000 R (Intelligenz-Struktur-Test 2000 R; IST 2000 R; Liepmann, Beauducel, Brocke, & Amthauer, 2007). This test is based on Thurstone's and Cattell's intelligence theories and measures verbal, numerical, and figural reasoning ability. The composite score indicates general reasoning ability, which is closely tied to general intelligence (*g*). The IST 2000 R is one of the most renowned and most applied intelligence tests in German-speaking countries. Its psychometric properties and validity are well established (e.g., Bühner, Ziegler, Krumm, & Schmidt-Atzert, 2006; also see Schmidt-Atzert & Rauch, 2008).

All students also reported their sex, age, and parents' educational levels (different from Study A, the answer "doctoral dissertation" was omitted so that only values from 0 to 16 were possible for the overall sum score). As in Study A, the schools provided the GPAs of the students from their last report cards before the transition to secondary level II.

3.1.3. Analyses

We again considered the ages of all G8 and G9 students as a covariate. Parental educational level did not differ between the two groups ($M_{\text{G8}} = 9.43$, $M_{\text{G9}} = 8.99$), $F(1, 446) = 1.94$, $p = 0.16$, $d = 0.13$.

We tested G8/G9 differences in broad intellectual abilities and in *g* by means of a higher-order multi-group analysis. In multi-group analyses, two approaches can be distinguished: the multi-group analysis of mean and covariance structures (MG-MACS) and multiple indicators-multiple causes (MIMIC) models. MIMIC models are more parsimonious and allow the inclusion of continuous covariates (e.g., Keith, Reynolds, Patel, & Ridley, 2008). Since we wanted to consider age as a covariate, we set up a MIMIC model with G8/G9 and age as exogenous variables predicting the latent cognitive abilities. G8/G9 was dummy-coded with G8 = 0 and G9 = 1. Therefore, positive path weights from G8/G9 to the latent abilities reflected differences in favor of the G9 students. For the evaluation of model fit, we used the χ^2 statistic, the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Usually, a CFI ≥ 0.97 is considered to indicate a very good model fit, and a CFI ≥ 0.95 to indicate a good or at least satisfactory fit; with regard to the RMSEA, values ≤ 0.05 and ≤ 0.08 indicate a good and an acceptable fit, respectively; for the SRMR, the thresholds for a good and an acceptable fit are ≤ 0.05 and ≤ 0.10, respectively (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

MIMIC models require stricter levels of invariance across groups than does MG-MACS. If not only configural, metric, and scalar invariance, but also invariance of measurement error, second-order factor loadings, and factor variances are met, results from MIMIC models equal those from MG-MACS (e.g., Keith, Reynolds, Patel, & Ridley, 2008). However, most levels of invariance cannot be tested in MIMIC models. Therefore, prior to conducting the MIMIC model, we performed MG-MACS to establish invariance by stepwise adding constraints for the respective model parameters (i.e., first-order factor loadings, subtest intercepts, residual variances, second-order factor loadings, factor variances). We used the χ^2 difference test as well as changes in CFI, RMSEA, and SRMR to evaluate whether model restriction led to a significant deterioration in fit. A ΔCFI ≥ -0.01, a ΔRMSEA ≥ 0.015 and a ΔSRMR ≥ 0.01 were taken as indicative of non-invariance (Chen, 2007). When scalar non-invariance occurred, we added a direct path in the MIMIC model from G8/G9 to the respective subtest to control for non-invariance caused by this subtest. All structural equation analyses were conducted with AMOS 24.0.

3.2. Results and summary

Before computing our main analyses, we again tested whether there were differences between the G8 and G9 students in their mastery of the curriculum contents. There was no statistically significant difference between the groups in their GPAs ($M_{\text{G8}} = 2.66$, $M_{\text{G9}} = 2.75$), $F(1, 432) = 2.66$, $p = 0.10$, $d = 0.16$.

First, we tested whether the data fitted the intelligence model the IST 2000 R is based on. For the G8 students, the fit was very good ($\chi^2 = 32.33$, $df = 24$, $p = 0.12$, CFI = 0.981, RMSEA = 0.038 [90% CI: 0.000; 0.068], SRMR = 0.042). For the G9 students, the fit was somewhat weaker but still satisfactory ($\chi^2 = 40.44$, $df = 24$, $p = 0.02$, CFI = 0.951, RMSEA = 0.058 [0.024; 0.088], SRMR = 0.057).

We then inspected factorial invariance by means of MG-MACS. Results are shown in Table 2. Both configural and metric

[3] The time gap between the assessments of the G8 and the G9 students was due to the availability of the data. It was not due to the implementation of the G8 reform or any other reasons associated with the G8 reform.

Table 2
Study B: Model Tests for Factorial Invariance.

Invariance Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	ΔCFI	RMSEA	ΔRMSEA	SRMR	ΔSRMR
Configural	72.77	48				0.968		0.034		0.042	
Metric	77.06	54	4.29	6	0.64	0.970	0.002	0.031	0.003	0.043	0.001
Scalar	152.39	60	75.33	6	< 0.001	0.881	−0.089	0.059	0.028	0.047	0.004
Partial scalar	81.98	58	4.92	4	0.30	0.969	−0.001	0.030	−0.001	0.040	−0.003
Measurement error	94.23	67	12.25	9	0.20	0.965	−0.004	0.030	< 0.001	0.041	0.001
Second-order factor loadings	100.47	69	6.24	2	0.04	0.960	−0.005	0.032	0.002	0.046	0.005
Factor variances	105.04	73	4.57	4	0.33	0.959	−0.001	0.031	−0.001	0.052	0.006

Note. N = 448. CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual. Cut-off values for (non)invariance: ΔCFI ≥ −0.01, ΔRMSEA ≥ 0.015, ΔSRMR ≥ 0.01 (Chen, 2007).

invariance were fully confirmed. Constraining the subtest intercepts, however, significantly reduced the model fit ($\Delta\chi^2 = 75.33$, $\Delta df = 6$, $p < 0.001$, ΔCFI = −0.089, ΔRMSEA = 0.028, ΔSRMR = 0.004). Inspection the modification indices (M.I.) showed that this lack of scalar invariance was due to the two subtests "Number Series" (M.I. = 15.63) and "Calculations" (M.I. = 11.67). When the intercepts of both subtests were freely estimated, the model fit did not deteriorate significantly ($\Delta\chi^2 = 4.92$, $\Delta df = 4$, $p = 0.30$, ΔCFI = −0.001, ΔRMSEA = −0.001, ΔSRMR = −0.003). Thus, at least partial scalar invariance was confirmed, and we proceeded with testing for the next steps of invariance needed for MIMIC models. Invariance of measurement errors was fully confirmed. When testing for invariance of second-order factor loadings, the χ^2 difference test was significant at $p = 0.04$. However, given that this statistic is relatively strict and no other index pointed to a lack of invariance (ΔCFI = −0.005, ΔRMSEA = 0.002, ΔSRMR = 0.005), we concluded that invariance of second-order factor loadings was given. Invariance of factor variances was also confirmed. Thus, the prerequisites for a meaningful interpretation of the results from a MIMIC model were fulfilled.

We first set up a MIMIC model without age as a covariate and then a MIMIC model with age as covariate so determine the effect age had on the results. Because the subtests "Number Series" and "Calculations" had shown unequal intercepts across the groups in MG-MACS, we added direct paths from G8/G9 to control for their scalar non-invariance.

Table 3 shows the results of both models. The model without age showed a very good fit to the data ($\chi^2 = 39.01$, $df = 28$, $p = 0.08$, CFI = 0.990, RMSEA = 0.030 [.000; 0.050], SRMR = 0.032). G8/G9 had a substantial effect (β = 0.39) on *g*, which equals a group difference of *d* = 0.85 in favor of the G9 students. The effects on the more specific abilities were smaller (verbal ability: β = 0.13, *d* = 0.26; numerical ability: β = 0.19, *d* = 0.39; figural ability: β = −0.06, *d* = −0.06). Thus, although G8/G9 was found to impact also relatively specific verbal and numerical abilities (when *g* was controlled for), the G8/G9 effect mainly related to *g*.

However, age was not considered in that analysis. Therefore, we ran the same model, but this time including age as a further exogenous variable (see Fig. 1). The fit of this model was reasonably good ($\chi^2 = 52.80$, $df = 34$, $p = 0.02$, CFI = 0.987, RMSEA = 0.035 [.014; 0.053], SRMR = 0.030). The effect on *g* was reduced to β = 0.34, which corresponds to a group difference of *d* = 0.72 in favor of the G9 students. Thus, comparable to Study A, about 85% (0.72 of 0.85) of the overall difference in *g* were due to the one year difference in school attendance, whereas about 15% were due to differences in age. The group differences in numerical and figural skills (i.e., when controlling for differences in *g*) remained largely unaffected when controlling for age. The group difference in specific verbal abilities increased to β = 0.22 (*d* = 0.45).

Thus, comparable to Study A, Study B indicated higher cognitive abilities of the G9 students as compared to the G8 students despite roughly the same mastery of the curriculum. Furthermore, structural equation modeling showed that the G8/G9 effect was primarily related to *g* (and not primarily to more specific skills unrelated to *g*).

4. General discussion

Prior research has repeatedly shown that schooling impacts intelligence test scores. However, the question of which abilities actually underlie the schooling effect has continued to be met with controversy. Two hypotheses have both appeared to be reasonable according to current research (see Ceci, 1991; Cliffordson & Gustafsson, 2008; Neisser et al., 1996): The schooling effect could be

Table 3
Study B: Effects of G8/G9 on Latent Cognitive Abilities with and without Chronological Age as Covariate.

	Effect of G8/G9 (without control for age)		Effect of G8/G9 (with control for age)	
	β	*d*	β	*d*
Verbal ability	0.13	0.26	0.22	0.45
Numerical ability	0.19	0.39	0.19	0.39
Figural ability	−0.06	−0.12	−0.05	−0.10
General reasoning ability	0.39	0.85	0.34	0.72

Note. The *d* values were calculated from the standardized path weights; positive *d* values and path weights indicate differences in favor of the G9 students.

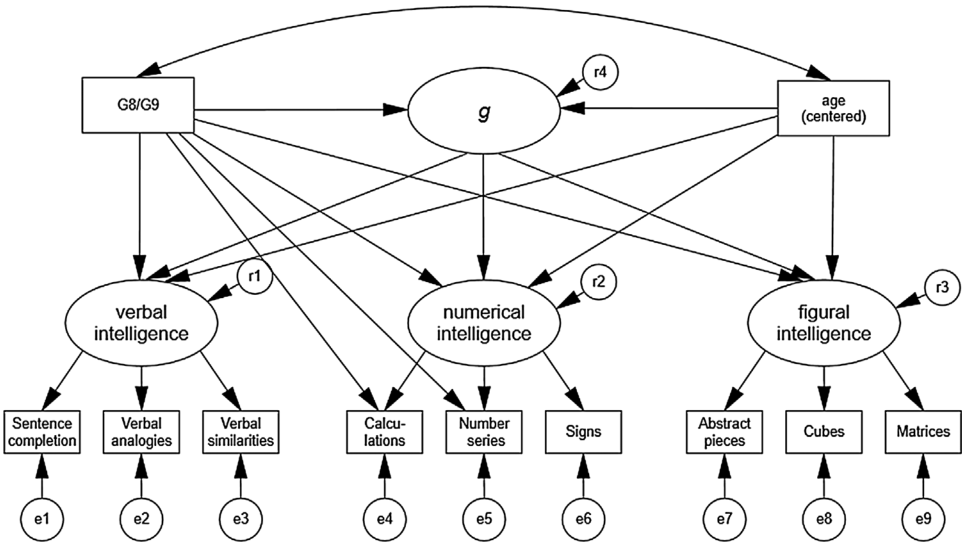

Fig. 1. Study B: MIMIC Model with G8/G9 Affiliation and Chronological Age as Exogenous Variables.

explained by either an enhancement of specific cognitive abilities that are closely linked to curricula or by an enhancement of more general, intelligence-related abilities that are mostly independent of curricula.

The German G8 school reform shortened the duration of school attendance but preserved the curriculum. Hence, we utilized the G8 reform as a natural quasi-experiment to test these two hypotheses against each other. If the effect of schooling on intelligence test scores were due to the fostering of specific abilities taught within the curricula, no difference was expected to emerge between the G8 and G9 students' intelligence test scores. On the other hand, if a difference were to emerge, it could be explained by an enhancement of intelligence-related abilities that are independent of the curriculum. We conducted two studies with different samples and different intelligence tests.

4.1. The nature of the effect of schooling on intelligence test scores

The results of the two studies were consistent and unambiguous. Despite completing the same curriculum, the G9 students outperformed the G8 students on nearly all of the cognitive abilities that were assessed. In Study A, the G9 students achieved statistically significantly higher factor scores of verbal, numerical, and figural cognitive contents as well as of operation speed and processing capacity and *g* as measured by the BIS test. In Study B, the G9 students outscored the G8 students on verbal, numerical, figural, and especially on general reasoning ability as measured by the IST 2000 R. Hence, in two different samples that were administered different intelligence tests, differences emerged between the G8 and G9 students' intelligence test scores in favor of the G9 students. The curricula for the G8 and G9 students were intended to be identical in both the number of lessons and the contents, and there were only very small differences in students' mastery of the curricular contents between the G8 and the G9 students as indicated by their GPAs. Thus, the differences in intelligence test scores could be primarily attributed to an enhancement of intelligence-related abilities by schooling, such as conditional reasoning (Artman & Cahan, 1993; Artman et al., 2006; Cahan & Artman, 1997; Cahan et al., 2008) or short-term memory (Morrison et al., 1995), rather than to an enhancement of specific cognitive skills and abilities that are tied to the curriculum.

How might schooling foster conditional reasoning or other intelligence-related abilities? There is still no clear answer to this question, but some assumptions can be made. In general, students are confronted with cognitively challenging tasks for many years in many different disciplines and are "trained" by different teachers (see also Rost, 2013). Students are encouraged to spend a lot of time and effort engaged in elaborated cognitive processes. It can be assumed that thinking is thereby successively disentangled from the concrete context of the problem and that a transfer of this thinking to more abstract levels is continuously supported and practiced over many years (see Vygotsky, 1978). Furthermore, schooling provides students not only with challenging tasks but also with experiences that allow them to disconfirm some rather naïve hypotheses that come from invalid conditional reasoning made in everyday life. For example, Artman et al. (2006) showed that extracurricular experiences had a slight negative effect on conditional reasoning, whereas schooling had a substantial positive effect. The conflict between invalid everyday life conclusions and the results of valid conclusions made and demonstrated in school seems to be beneficial for intellectual development as such conflict can lead to cognitive accommodation (see Artman et al., 2006; Cahan & Artman, 1997; Cahan et al., 2008).

Solid studies that have investigated the extent to which intelligence can be fostered by relatively brief trainings have mostly shown disappointing results (see Rost, 2013, pp. 417–430, for a review). It seems to be the case that a sustained promotion of intelligence needs many years of intensive, continuous training. Obviously, every (additional) year of schooling seems to contribute to intellectual development. It can be assumed that every additional year of promotion helps to anchor cognitive abilities. Moreover, every year of schooling could offer more opportunities to apply and to practice these abilities, a process that in turn should stabilize them further.

4.2. Size of the effects

The effects in Study A were small to medium in size. The size of the effect on *g* was 7.6 IQ points, which replicates the effect on *g* found by Stelzl et al. (1995) and exceeds the effect reported in Rindermann's review (2011) by only 2 IQ points. The effect size for *g* in Study B was 10.8 IQ points. This effect exceeds the size that would be expected on the basis of one year of schooling as indicated by previous research. This is probably due to the fact that the G9 students in Study B were tested 5 years later than the G8 students in Study B, which has two important implications. First, in the last decade the quota of students in the secondary school system attending the Gymnasium has increased. This increase was not very strong between 2007/2008 and 2012/2013 (i.e., when we tested the students): in North Rhine-Westphalia, the attendance rate increased from 42.2% to 44.9% (Statistisches Bundesamt, 2009, 2014). However, it might be that there were some more students with weaker cognitive abilities in the G8 cohort than in the G9 cohort from the outset. Second, according to the first results from some European countries, the Flynn effect may have been reversed since the late 1990s by a decrease of about 2.5 IQ points per decade (e.g., Dutton & Lynn, 2013, 2015; Teasdale & Owen, 2008; see Dutton, van der Linden, & Lynn, 2016, for review). Given the time interval of 5 years between the testing of the G8 and G9 cohorts in Study B, a small part of the difference in intelligence in favor of the G9 students might also be explained by a reversal of the Flynn effect. The reasons for this possible reversal are still a subject of debate (see, e.g., Dutton et al., 2016). However, what perhaps might have contributed to a possible decrease in IQ within the 5-year gap unrelated to schooling are contextual factors which might have become especially pertinent in this time interval. Nevertheless, although we could not control for such factors, it appears unlikely that they account for the complete effect, given that the results from Study A without any time lag were comparable.

Some previous studies found weaker effects of schooling on intelligence test scores (Cliffordson & Gustafsson, 2008; Rost & Wild, 1995). Therefore, our effect sizes are among the upper bound of the effect sizes found in previous studies. It is important to note, however, that most of these studies were based on samples that were not highly selective, such as elementary school children. Yet, students in German Gymnasiums are known to be highly selected for their intelligence (e.g., Steinmayr, Beauducel, & Spinath, 2010; Steinmayr, Bergold, Margraf-Stiksrud, & Freund, 2015). This typically leads to a variance restriction in intelligence in this student population. Thus, the *d*-values in our study have probably provided at least slight overestimations of the "true" effect.

4.3. Limitations

The G8 reform was utilized as a "natural" experiment so that there was no possibility to randomize students to the G8 or G9 condition. However, there was no self-selection of students because students or their parents were not allowed to choose between G8 and G9. Unfortunately, we were not able to test the students' intelligence when they first entered the Gymnasium, and hence, we could not control for their base level intelligence test scores. This is problematic for Study B in which the G8 cohorts were tested 5 years later than the G9 cohorts (see above). In Study A, however, the degree of group-specific selection was most likely not a problem. The time interval between the assessments of both cohorts was only 1 year. The two cohorts were equal in both size and distribution of sex and comparable on their parents' educational levels and their GPAs. Furthermore, the G8 reform was simultaneously implemented for *all* Gymnasiums in the entire federal state of North Rhine-Westphalia. Thus, our participants were not part of a model test for G8 wherein parents, students, or teachers could decide whether the students would follow the G8 or G9 format. There were no other Gymnasiums in the state that did not participate in the G8 implementation. Thus, there were no alternatives for families that did not want their children to be taught as required by the G8 reform.

There was no other way for us to account for chronological age but to hold it statistically constant. This was not an optimal strategy because there was only a small overlap in age between the G8 and G9 students. In other words, G8/G9 affiliation and age were substantially confounded (Study A: $r = 0.82$; Study B: $r = 0.78$), posing the problem of collinearity (even though we could account for the high correlation between G8/G9 and age at least in Study B, using a structural equation model where we allowed both variables to intercorrelate). Therefore, a satisfying equation on age could not be implemented. However, previous research has consistently found that from elementary school age onwards, the effect of age on intelligence is relatively small, all the more when directly compared with the effect of schooling (Cahan & Cohen, 1989; Cliffordson & Gustafsson, 2008; see also Hein, Tan, Aljughaiman, & Grigorenko, 2015). Also studies based on German student samples have found that 1 year of chronological age makes only a small contribution to intellectual development, as opposed to schooling, which has strong effects (Merz et al., 1985; Rost & Wild, 1995; Stelzl et al., 1995; see 1.1). The weakness of the age effect is further underlined by the fact that the standardization of both the BIS test and the IST 2000 R does not differentiate between 15- and 16-year-olds (i.e., the age groups we investigated).

Our conclusions rest on the fact that both the curriculum contents and amount of hours spent at school were identical for the G8 and G9 students. In line with this fact, we found no statistically significant differences in the G8 and the G9 students' mastery of the contents of the curriculum as indicated by their GPAs. In Study A, the descriptive data showed a small difference in the grasp of the curriculum in favor of the G9 students. Therefore, one might still argue that even though the curriculum was equal between groups, the G9 students in Study A might have used their additional year of time to practice and to apply the contents in their daily lives somewhat more than the G8 students. Therefore, curriculum-tied skills might still have added to the enhancement of intelligence test scores in Study A. However, in Study B we found an effect of schooling on intelligence test scores although the G9 students had slightly worse GPAs than the G8 students.

A further point that deserves discussion is the question whether our findings might alternatively be explained by negative effects of the curriculum acceleration the G8 students underwent. The G8 students had to complete the same curriculum in one year less time, which was done by extending the lessons into the afternoon. One might argue that this acceleration might have been detrimental for the G8 students' intellectual development, and that this could have caused our findings. According to our design, we cannot fully rule out this possibility. However, acceleration is not necessarily disadvantageous for students' cognitive development. Acceleration in instruction has rather been found to promote children's and adolescents' academic performance, not only in gifted but in most students (see the extensive meta-analysis by Hattie, 2009).

As a final limitation, we tested students from only two schools. Hence, we cannot definitely rule out the possibility that the differences found in intelligence might also be due to some special characteristics of these schools. For example, one might argue that some schools might not have been fully prepared for the G8 reform and were not able to provide the G8 students with the same amenities like the G9 students. However, first, the G8 and the G9 students were comparable on their GPAs, and, second, the teaching materials for the G8 students were the same as for the G9 students. Furthermore, the schools in our sample did not differ in any obvious ways from other Gymnasiums in the federal state where we conducted our study. Therefore, this alternative explanation appears rather unlikely.

4.4. Conclusions

There is still an ongoing debate about which processes underlie the effect of schooling on intelligence test scores. The G8 reform provided us with the opportunity to shed some light on this matter, that is, to test whether the schooling effect is driven by either the teaching of specific, curriculum-tied abilities or the fostering of rather general, curriculum-independent abilities that are not explicitly taught in school. Our study suffers from some limitations that are not unusual in "natural" (quasi-)experiments. Therefore, our results have to be interpreted with caution. However, we were able to conduct two independent studies with different samples and different intelligence tests and to replicate our findings, thus providing support for the validity of our conclusions. We consistently found evidence against the hypothesis that the effect of schooling on intelligence is solely due to the teaching of cognitive skills that are tied to the curriculum. Rather, our results supported the hypothesis that the effect of schooling on intelligence is mainly due to the fostering of intelligence-related abilities that are irrespective of formal curricula. This finding supports the idea that the enhancement of intelligence test scores due to schooling is mainly the consequence of a "real" enhancement of intelligence and not just a consequence of teaching students "a reservoir of IQ-relevant knowledge and [shaping] their style of responding" (Ceci, 1991, p. 717), leading to an artificial enhancement of intelligence test scores. Additional studies need to be conducted to further replicate our findings. However, our study can be viewed as one step toward clarifying the nature of the effect of schooling on intelligence test scores.

References

Alexander, J. R. M., & Martin, F. (2004). The end of the reading age: Grade and age effects in early schooling. *Journal of School Psychology, 42*, 403–416.

Amin, V., Behrman, J. R., & Spector, T. D. (2013). Does more schooling improve health outcomes and health related behaviors? Evidence from U.K. twins. *Economics of Education Review, 35*, 134–148.

Artman, L., & Cahan, S. (1993). Schooling and the development of transitive inference. *Developmental Psychology, 29*, 753–759.

Artman, L., Cahan, S., & Avni-Babad, D. (2006). Age, schooling and conditional reasoning. *Cognitive Development, 21*, 131–145.

Bühner, M., Ziegler, M., Krumm, S., & Schmidt-Atzert, L. (2006). Ist der I-S-T 2000 R Rasch-skalierbar? [Is the I-S-T 2000 R Rasch scaleable?]. *Diagnostica, 52*, 119–130.

Baker, D. P., Salinas, D., & Eslinger, P. J. (2012). An envisioned bridge: Schooling as a neurocognitive developmental institution. *Developmental Cognitive Neuroscience*, 6–17.

Baltes, P. B., & Reinert, G. (1969). Cohort effects in cognitive development of children as revealed by cross-sectional sequences. *Developmental Psychology, 1*, 169–177.

Beckmann, J. F., & Guthke, J. (1999). Testinformation. Jäger, A.O., Süß, H.-M. & Beauducel, A. (1997). Berliner Intelligenzstruktur-Test (BIS), Form 4 [Test information. Jäger, A.O., Süß, H.-M., & Beauducel, A. (1997). Berlin Intelligence Structure Test (BIS), Form 4]. *Diagnostica, 45*, 56–61.

Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology, 31*, 221–236.

Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence, 33*, 93–106.

Bowles, S., & Gintis, H. (2002). Schooling in capitalist America revisited. *Sociology of Education, 75*, 1–18.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 425–430.

Cahan, S., & Artman, L. (1997). Is everyday experience dysfunctional for the development of conditional reasoning? *Cognitive Development, 12*, 261–279.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*, 1239–1249.

Cahan, S., Greenbaum, C., Artman, L., Deluya, N., & Gappel-Gilon, Y. (2008). The differential effects of age and first grade schooling on the development of infralogical and logico-mathematical concrete operations. *Cognitive Development, 23*, 258–277.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.

Cliffordson, C., & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation of age and length of schooling. *Intelligence, 36*, 143–152.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology, 91*, 604–614.

Cunningham, A., & Carroll, J. (2011). Age and schooling effects on early literacy and phoneme awareness. *Journal of Experimental Child Psychology, 109*, 248–255.

Dutton, E., & Lynn, R. (2013). A negative Flynn effect in Finland, 1997–2009. *Intelligence, 41*, 817–820.

Dutton, E., & Lynn, R. (2015). A negative Flynn effect in France, 1999 to 2008–9. *Intelligence, 51*, 67–70.

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn effect: A systematic literature review. *Intelligence, 59*, 163–169.

Gustafsson, J.-E. (2008). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. In P. C. Kyllonen, R. D. Roberts, & L. Stankov (Eds.). *Extending intelligence. Enhancement and new constructs* (pp. 31–50). New York, NY: Lawrence Erlbaum.

Hansen, K. T., Heckman, J. J., & Mullen, K. T. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics, 121*, 39–98.

Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* London, UK: Routledge.

Hein, S., Tan, M., Aljughaiman, A., & Grigorenko, E. L. (2015). Gender differences and school influences with respect to three indicators of general intelligence: Evidence from Saudi Arabia. *Journal of Educational Psychology, 107*, 486–501.
Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4. [Berlin Intelligence Structure Test. Form 4].* Göttingen, Germany: Hogrefe.
Jäger, A. O. (1984). Intelligenzforschung. Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Research on the structure of intelligence. Competitive models, new developments perspectives]. *Psychologische Rundschau, 35*, 21–35.
Kühn, S. M., van Ackeren, I., Bellenberg, G., Reintjes, C., & im Brahm, G. (2013). Wie viele Schuljahre bis zum Abitur? Eine multiperspektivische Standortbestimmung im Kontext der aktuellen Schulzeitdebatte [How many years until abitur in German upper secondary schooling? Taking stock in the context of current school duration debates]. *Zeitschrift für Erziehungswissenschaft, 16*, 115–136.
Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence, 36*, 502–525.
Kemptner, D., Jürges, H., & Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics, 30*, 340–354.
Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (2., erweiterte und überarbeitete Aufl.) [Intelligence-Structure-Test 2000 R (2nd, extended and revised ed.)].* Göttingen, Germany: Hogrefe.
Merz, F., Remer, H., & Ehlers, T. (1985). Der Einfluß des Schulbesuchs auf Intelligenzleistungen im Grundschulalter [Schooling's impact on intellectual achievement in elementary school age]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 17*, 223–241.
Milde-Busch, A., Blaschek, A., Borggräfe, I., von Kries, R., Straube, A., & Heinen, F. (2010). Besteht ein Zusammenhang zwischen der verkürzten Gymnasialzeit und Kopfschmerzen und gesundheitlichen Belastungen bei Schülern im Jugendalter? [Is there an association between the reduced school years in grammar schools and headache and other health complaints in adolescent students?]. *Klinische Pädiatrie, 222*, 255–260.
Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology, 31*, 789–799.
Naito, M., & Miura, H. (2001). Japanese children's numerical competencies: Age- and schooling-related influences on the development of number concepts and addition skills. *Developmental Psychology, 37*, 217–230.
Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.
Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence, 36*, 236–260.
Rindermann, H. (2011). Intelligenzwachstum in Kindheit und Jugend [Increase of intelligence in childhood and youth]. *Psychologie in Erziehung und Unterricht, 58*, 210–224.
Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology, 51*, 573–582.
Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child Development, 86*, 23–36.
Rost, D. H., & Wild, K.-P. (1995). Fördert die Schule die Intelligenz? [Does schooling foster intelligence?]. *Paper presented at the 51st Meeting of the Research Group for Empirical Educational Research (AEPF)*.
Rost, D. H. (1993). *Lebensumweltanalyse hochbegabter Kinder [Life-environmental analysis of gifted children].* Göttingen, Germany: Hogrefe.
Rost, D. H. (2013). *Handbuch Intelligenz [Handbook of intelligence].* Weinheim, Germany: Beltz.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.
Schmidt-Atzert, L., & Rauch, W. (2008). TBS-TK Rezension. Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage [TBS-TK review. Intelligence-Structure-Test 2000 R (I-S-T 2000 R). 2nd, extended and revised edition]. *Report Psychologie, 33*, 303–304.
Schnittker, J., & Behrman, J. R. (2012). Learning to do well or learning to do good? Estimating the effects of schooling on civic engagement, social cohesion, and labor market outcomes in the presence of endowments. *Social Science Research, 41*, 306–320.
Scribner, S., & Cole, M. (1973). Cognitive consequences of formal and informal education. *Science, 182*, 553–559.
Standing Conference of Ministers of Education and Cultural Affairs for the Federal States in the Federal Republic of Germany (1968). *Erläuterung der Notenstufen bei Schulzeugnissen und Einzelergebnissen in staatlichen Prüfungszeugnissen [Explanation of the grades in report cards and individual results in state examination certificates]* Resolution No. 675, October 3.
Statistisches Bundesamt (Ed.). (2009). *Statistisches Jahrbuch 2009. Für die Bundesrepublik Deutschland [Statistical yearbook 2009. For the Federal Republic of Germany].* Wiesbaden, Germany: Statistisches Bundesamt.
Statistisches Bundesamt (Ed.). (2010). *Statistisches Jahrbuch 2010. Für die Bundesrepublik Deutschland mit „Internationalen Übersichten [Statistical yearbook 2010. For the Federal Republic of Germany with „international schemata].* Wiesbaden, Germany: Statistisches Bundesamt.
Statistisches Bundesamt (Ed.). (2012). *Statistisches Jahrbuch. Deutschland und Internationales. 2012 [Statistical yearbook. Germany, and international affairs. 2012].* Wiesbaden, Germany: Statistisches Bundesamt.
Statistisches Bundesamt (Ed.). (2014). *Statistisches Jahrbuch. Deutschland und Internationales. 2014 [Statistical yearbook. Germany, and international affairs. 2014].* Wiesbaden, Germany: Statistisches Bundesamt.
Statistisches Bundesamt (Ed.). (2015). *Statistisches Jahrbuch. Deutschland und Internationales. 2015 [Statistical yearbook. Germany, and international affairs. 2015].* Wiesbaden, Germany: Statistisches Bundesamt.
Steinmayr, R., Bergold, S., Margraf-Stiksrud, J., & Freund, P. A. (2015). Gender differences on general knowledge tests: Are they due to differential item functioning? *Intelligence, 50*, 164–174.
Steinmayr, R., Beauducel, A., & Spinath, B. (2010). Do sex differences in a faceted model of fluid and crystallized intelligence depend on the method applied? *Intelligence, 38*, 101–110.
Steinmayr, R., Dinger, F. C., & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. *European Journal of Personality, 24*, 535–550.
Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and crystallized intelligence: A quasi-experimental study. *Intelligence, 21*, 279–296.
Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence, 36*, 121–126.
Trautwein, U., Hübner, N., Wagner, W., & Kramer, J. (2015). *Konsequenzen der G8-Reform–Zusammenfassung zentraler Befunde [Consequences of the G8 reform–Summary of the main findings].* Tübingen, Germany: Eberhard Karls Universität, Hector Institut für Empirische Bildungsforschung.
Van de Vijver, F. J. R., & Brouwers, S. A. (2009). Schooling and basic aspects of intelligence: A natural quasi-experiment in Malawi. *Journal of Applied Developmental Psychology, 30*, 67–74.
Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes.* Cambridge, MA: Harvard University Press.