



# Inclusive Education of Students With General Learning Difficulties: A Meta-Analysis

Sonja Krämer , Jens Möller,  
and Friederike Zimmermann 

*Institute for Psychology of Learning and Instruction, Kiel University*

*This article presents a meta-analysis on cognitive (e.g., academic performance) and psychosocial outcomes (e.g., self-concept, well-being) among students with general learning difficulties and their peers without learning difficulties in inclusive versus segregated educational settings. In total, we meta-analyzed  $k = 40$  studies with 428 effect sizes and a total sample of  $N = 11,987$  students. We found a significant small to medium positive effect for cognitive outcomes of students with general learning difficulties in inclusive versus segregated settings ( $d = 0.35$ ) and no effect on psychosocial outcomes ( $d = 0.00$ ). Students without general learning difficulties did not differ cognitively ( $d = -0.14$ ) or psychosocially ( $d = 0.06$ ) from their counterparts in segregated settings. We examined several moderators (e.g., design, diagnosis, type of outcome). We discuss possible selection effects as well as implications for future research and practice.*

**KEYWORDS:** meta-analysis, inclusive education, learning difficulties, cognitive outcomes, psychosocial outcomes

In 2006, the United Nations passed the Convention on the Rights of Persons with Disabilities (CRPD; United Nations, 2006), which stated that persons with disabilities and corresponding special educational needs (SEN) should have the opportunity to be educated in the general educational system and should not be excluded because of their disabilities (see Article 24). This established a human right to participation for people with SEN and emphasized that all people must be treated equally, regardless of their individual characteristics. The CRPD served as an impetus for many countries to create an inclusive educational system in which students with and without SEN are taught together. Whereas students with SEN were previously predominantly taught in segregated settings, the proportion of students with SEN in inclusive settings has increased in recent years. In the United

States, the percentage of students with SEN in inclusive education increased by 16% within 17 years (McFarland et al., 2019), reaching an inclusion rate of 63% in 2017. Europe has a comparable share of students with SEN in inclusive education, at about 61% in 2014/2015, representing an increase of about 8% in only 2 years (53% in the 2012/2013 school year; European Agency for Special Needs and Inclusive Education, 2017, 2018). These efforts toward the development of an inclusive school system have been linked to a new view on students' heterogeneity. While students previously tended to be divided into distinct groups (e.g., students with intellectual disabilities vs. students without intellectual disabilities), learning-related challenges are now more likely to be seen as continuously distributed (e.g., Craddock & Owen, 2005; Feczko et al., 2019). This suggests that students do not fundamentally differ from one another but have different support needs. In summary, differences between students are increasingly viewed as more quantitative than qualitative. This is often accompanied by the demand that all students be offered the same educational opportunities through an inclusive school system because they deserve an equitable education alongside their peers.

Despite the increasing number of students with SEN enrolled in inclusive education, many students are still taught in segregated educational settings. Therefore, an open question concerns in which educational setting (inclusive or segregated) students with SEN experience better outcomes. Moreover, the inclusion of students with SEN could also have an impact on students without SEN—for example, increasing achievement heterogeneity as a result of changes in class composition. Thus, another open question concerns whether students without SEN benefit more from being taught together with students with SEN in inclusive educational settings or being taught separately. Previous meta-analyses indicated predominantly positive to neutral effects of including students with SEN in inclusive classes (Carlberg & Kavale, 1980; Oh-Young & Filler, 2015; Szumski et al., 2017; Wang & Baker, 1985). These results concern both cognitive (e.g., academic performance) and psychosocial outcomes (e.g., self-concept, anxiety, wellbeing) among both students with SEN and their peers without SEN.

Nevertheless, previous meta-analyses focus mainly on *all* students with any kind of SEN. Instead, it can be assumed that the effects of inclusive education differ depending on the type and extent of a student's SEN (see Cooc, 2019). For example, Carlberg and Kavale (1980) showed in their older meta-analysis that both students with IQs from 50 to 75 and those with IQs from 75 to 90 in inclusive classes outperformed their counterparts in segregated settings. The students with less severe disabilities (IQ 75–90) particularly benefited from inclusive education. However, students with specific limitations in learning *and* emotional and behavioral disorders exhibited poorer academic performance and less positive social outcomes in inclusive settings than in segregated settings. It also probably plays a role for peers which SEN their classmates in inclusive settings have. For example, the inclusion of students with emotional and behavioral disorders can have more negative effects on students without SEN than other common SEN such as learning disabilities (e.g., Fletcher, 2010; Friesen et al., 2010; Kuhl et al., 2020). A reason for this might be that students with emotional and behavioral disorders are more likely to disrupt teaching and therefore learning processes through externalizing behaviors than students with other disabilities (cf. Becherer

et al., 2020). Therefore, a differential consideration of the effects of different kinds of SEN is necessary.

The Organisation for Economic Co-operation and Development (OECD; 2007) has compiled an overview of definitions of different types of SEN in different countries. One of the largest groups of students receiving special educational support in either inclusive or special education settings are students with general limitations in learning across several school subjects (e.g., Banks & McCoy, 2011). These students must be distinguished from students with specific limitations in learning in individual school subjects that do not affect their performance in multiple subjects. Furthermore, students with more extensive cognitive limitations that affect various aspects in daily life and are not primarily limited to learning must be differentiated from students with (more mild) general limitations in learning. These three types of SEN with respect to limitations in learning are differentiated on a conceptual level in numerous countries (OECD, 2007).

However, there is great variability regarding the terminology used to describe these types of SEN between and even within countries. For example, terms that are frequently used for the above-mentioned types of SEN are *learning disability*, *mental retardation*, *intellectual disability*, or *learning difference* (see, e.g., Learning Disability Association of New York, n.d.; OECD, 2007; Schalock et al., 2007; World Health Organization, 2004). All of these terms can include various types and severities of learning limitations. Moreover, the boundaries between different types and severities are fluid, adding to the fuzziness of the distinctions between terms. Nevertheless, it is important to differentiate between some basic forms and degrees of severity of learning limitations, since the effects of inclusion can be expected to be quite different for these groups of students (e.g., Bakker et al., 2007). To clearly demarcate the target group of interest in this meta-analysis, which consists exclusively of students with more mild general limitations in learning, we use the term *general learning difficulties* (GLD).

Learning difficulties are defined as students' difficulties in performing academically at school at a level deemed appropriate for their age group (e.g., U.K. Public General Acts, 2014, Section 20) or as an accordingly "severe, extensive and long-lasting deficiency in their learning capacity" (OECD, 2007, p. 54). Thus, the term *general learning difficulties* clearly focuses on students with general difficulties in learning that affect their performance in almost all school subjects. A mildly below-average IQ of about 50 to 90 is frequently used as an additional diagnostic criterion for GLD beyond general limitations in learning in various subjects, but the exact IQ thresholds differ between countries (see OECD, 2007). The IQ ranges often used to diagnose GLD can be classified according to ICD-10 (International Statistical Classification of Diseases and Related Health Problems—10th Revision; World Health Organization, 2004), with an IQ between 71 and 84 referred to as *borderline intellectual functioning* (coded as R41.83), while an IQ between 55 and 70 is termed a *mild intellectual disability* (coded as F70). Students with diagnosed GLD often fall into one of these categories in practice.

By using this definition of GLD, we exclude students with specific limitations in learning and students with more severe limitations in learning as well as multiple domains of daily life. A differentiation between GLD and specific limitations in learning is useful because these students' learning occurs under different

conditions. Students with GLD usually have intellectual impairments due to lower general cognitive abilities. In contrast, poor achievement by students with specific limitations in learning cannot be explained by basic intellectual impairments, as they usually have an average IQ, which is one of the diagnostic criteria in the ICD-10 (coded as F81.0-3). A distinction is also made between students with GLD and students with more severe intellectual disabilities. The difference is the extent of impairment: Compared to students with GLD, students with more severe intellectual disabilities often have a lower IQ and not only are impaired in learning but also have difficulties in several domains of daily life, such as communication skills (Schalock et al., 2007).

In this meta-analysis, we provide a comprehensive overview of the effects of the inclusive schooling of students with GLD as defined above. Our understanding of inclusion is based on the CRPD and thus involves students with and without disabilities being taught together in a general educational system (United Nations, 2006). The aim of this meta-analysis is to evaluate the effects of inclusion based on the general placement definition. On one hand, we examine the effects of an inclusive educational system on students with GLD themselves. Not only their academic performance is relevant but also psychosocial factors such as their self-concepts, feeling of social integration, or possible (test) anxieties. Therefore, we consider both cognitive and psychosocial outcomes. On the other hand, we examine to what extent students without GLD are affected by the inclusion of students with GLD. The group of students without GLD includes all classmates who have no SEN of any type. We also considered both cognitive and psychosocial outcomes for these students.

## **Theoretical and Empirical Background**

### *Theoretical Assumptions*

In this section, we consider possible opportunities and challenges in terms of student outcomes for inclusive education compared to segregated education from a theoretical point of view. Segregated educational settings refer to the practice of enrolling students with GLD in special schools and students without GLD in regular schools. In order to simplify the terminology, both types of schools will be referred to here as segregated educational settings. Why might students with GLD and their peers without GLD benefit from inclusive education? And why might inclusion be detrimental for both groups? To answer these questions, this section presents the potential pros and cons for students with and without GLD, starting with cognitive outcomes and continuing on to psychosocial variables.

### *Students With GLD*

In terms of *cognitive outcomes*, composition effects could contribute to better academic performance in inclusive settings. Compositional effects indicate that the development of students' achievement depends on the composition of the learning group (Coleman et al., 1966; Thrupp et al., 2002). In particular, the peer effect, as one composition effect, states that students perform better when placed in classes with higher performing peers (e.g., Harker & Tymms, 2004; Justice et al., 2014). Since the class average achievement level in inclusive classes tends

to be higher than in segregated education settings for students with GLD, such students might benefit cognitively from being placed in inclusive classes compared to lower performing groups in segregated settings. Potential reasons for this could be that students with LD orient themselves toward students without GLD and see them as role models (cf. Kocaj et al., 2014). They may adopt successful learning strategies from students without GLD and thus perform better academically (e.g., Slavin, 1996). Furthermore, a higher average academic performance level in inclusive classes might lead to teachers having higher performance expectations of their students (e.g., Dar & Resh, 1986; Hornstra et al., 2010). This could lead teachers to give students more challenging tasks (e.g., Diamond et al., 2004; Markussen, 2004). Tasks that are slightly above students' actual performance level often lead to an increase in performance (e.g., Hattie, 2009).

However, inclusive education can also have disadvantages for cognitive outcomes among students with GLD. First, the higher performance expectations and challenging tasks mentioned above might also harm students' performance: If these are not in alignment with students' actual performance levels, they can become overwhelmed. This overburdening can be perceived by students as academic failure, which can in turn lead to demotivation and frustration (e.g., Daniel & King, 1997), which then manifests in poorer performance (e.g., Graham & Weiner, 1996; Linnenbrink & Pintrich, 2002). Second, inclusive classes tend to be more heterogeneous in terms of students' academic performance than segregated settings. This could lead teachers to be less responsive to their students' individual performance levels and to a tendency to teach in an undifferentiated way, mainly for students with close to average performance. A potential consequence of this is that low-performing students might be overwhelmed (cf. Cole et al., 2004). Third, inclusive classes are usually larger than segregated classes for students with GLD (Hocutt, 1996). This may hinder teachers' ability to concentrate on individual students with GLD in inclusive classes due to less free capacity (cf. Staub & Peck, 1994). Less individual support and in turn worse academic outcomes for students with GLD can be the result. Fourth, if students with GLD perceive themselves as inferior compared to their classmates, it could lead to a low academic self-concept (Möller et al., 2009) and demotivation and thus also poorer performance (e.g., Allodi, 2000).

Beyond these advantages and disadvantages of inclusive education, it is also important to consider a possible selection effect. Students with GLD might differ regarding their cognitive abilities even before they enter an inclusive or segregated educational setting: Students with GLD with comparatively higher cognitive abilities might more frequently enroll in inclusive settings, while their counterparts with lower abilities enroll more often in segregated settings (e.g., Dessemontet et al., 2012; Madden & Slavin, 1983; Möller, 2013). This could explain differences in performance between students with GLD in inclusive versus segregated schools that cannot be attributed to the type of educational setting.

The literature discusses not only cognitive effects of inclusive schooling but also *psychosocial outcomes*, such as self-concept, social integration, anxiety, or well-being. On the one hand, inclusive schooling could have positive effects on students' self-concepts and self-esteem in the sense of a "basking in reflected

glory” effect (BIRG; e.g., Cialdini et al., 1976; Vogl & Preckel, 2014). The BIRG effect states that students identify themselves with the successes of the class and their classmates, increasing their own self-esteem. Furthermore, students with GLD in inclusive settings can get to know more diverse peers and might benefit in terms of their feeling of social integration (e.g., Nakken & Pijl, 2002; Ruijs & Peetsma, 2009). A reason for this could be that there are more regular schools with inclusive classes than segregated schools for students with GLD, meaning that students in inclusive settings are often able to attend school closer to where they live (e.g., Daniel & King, 1997). This may lead to students with GLD also being better integrated outside of school through a circle of school friends who live in their neighborhood.

On the other hand, it is assumed that students with GLD in inclusive settings might be disadvantaged in terms of psychosocial outcomes compared to students with GLD in segregated settings. In inclusive settings, a “big fish little pond” effect (BFLPE; Marsh et al., 2019) concerning students’ self-concepts may occur. According to the BFLPE, students in high-performing classes develop a worse academic self-concept than students with equal individual performance in low-performing classes. Students with GLD in inclusive settings compare their performance to that of their classmates without GLD. Because their performance is below the class average, they develop lower academic self-concepts. In segregated schools for students with GLD, the average performance level is much lower, so students with GLD make less frustrating upward comparisons and thus develop less negative self-concepts than their counterparts in inclusive settings (cf. Gresham & MacMillan, 1997).

Furthermore, students with GLD may be socially excluded in inclusive settings because they differ in various ways from students without GLD—besides intellectual capability, also in their mood (e.g., Lackaye et al., 2006), for example, as well as in emotional aspects (e.g., Gallegos et al., 2012). Moreover, because they lack a protected environment, as would be the case in segregated settings, students with GLD may suffer from higher pressure to perform well, feel frustrated, and thus develop school-related anxieties (e.g., Bear et al., 2002).

### *Students Without GLD*

The effects of inclusive education on cognitive and psychosocial outcomes among students without GLD should also be considered. Students without GLD in inclusive settings may benefit in terms of *cognitive outcomes* from more adaptive lessons due to additional teaching staff (e.g., Dyson et al., 2004). However, their cognitive outcomes might be negatively affected by teachers paying less attention to students without GLD and a lower average achievement level in the class (Huber et al., 2001; Staub & Peck, 1994).

Unlike students with GLD, it is possible that students without GLD may experience disadvantages due to composition effects (Thrupp et al., 2002). The inclusion of students with GLD reduces the average performance level of the class, which may result in students without GLD performing worse compared with students without GLD in noninclusive settings. In contrast to the advantages of composition effects for students with GLD, students without GLD in inclusive settings may follow their peers’ less successful learning strategies. Furthermore, students

without GLD might feel unchallenged by a lower average performance level in class, causing them to become demotivated and thus perform worse themselves (e.g., Linnenbrink & Pintrich, 2002).

With regard to *psychosocial outcomes*, students without GLD in inclusive settings may develop less fear of contact and prejudices as well as more positive attitudes toward students with GLD (e.g., De Boer et al., 2012). The contact hypothesis suggests a potential reason for this (e.g., Allport, 1954; Keith et al., 2015). Furthermore, students without GLD can gain social skills in interacting with people with disabilities (e.g., Ogelman & Seçer, 2012). According to the BFLPE, the lower average performance level in the class may lead to more downward comparisons and result in a higher academic self-concept among students without GLD in inclusive settings.

### *Empirical Findings*

Scholars assume that inclusive education might have both positive and negative effects on students with GLD and their peers without GLD. Empirical studies have examined the concrete effects of inclusive education.

### *Students With GLD*

To our knowledge, there are no recent reviews or meta-analyses specifically focused on the outcomes of students with GLD in inclusive educational settings. The most recent meta-analysis on the effects of inclusive education referred to students with any kind of SEN (Oh-Young & Filler, 2015). In the 1980s, two meta-analyses investigated the effect of inclusive education on students with SEN in general and also calculated effect sizes for different subgroups of SEN (Carlberg & Kavale, 1980; Wang & Baker, 1985).

Carlberg and Kavale (1980) examined mean effect sizes for three different subgroups of students with SEN: students with an IQ between 50 and 75, students with an IQ between 75 and 90, and learning-disabled students with emotional and behavioral disorders. The results indicated that students with IQs from 50 to 75 and IQs from 75 to 90 exhibited more positive cognitive and psychosocial outcomes in inclusive classes than in segregated classes ( $d = 0.14$  or  $d = 0.34$ , respectively), while learning-disabled students with emotional and behavioral disorder experienced more positive outcomes in segregated than in inclusive classes ( $d = -0.29$ ). Wang and Baker (1985) showed that intellectually disabled students in inclusive settings outperformed their counterparts in segregated education in performance ( $d = 0.43$ ) and process outcomes (e.g., type of interactions between teachers and students;  $d = 0.55$ ) but not in attitudinal outcomes (e.g., self-concept, attitudes toward learning;  $d = 0.01$ ).

Elbaum (2002) conducted a meta-analysis on the effects of different educational settings on the self-concept of students with GLD. He compared the self-concepts of students with GLD along a continuum from less restrictive educational settings (e.g., regular classroom for all instruction) to more restrictive settings (e.g., special education schools) and largely found no overall differences. Only the self-concept of students in self-contained classes within regular schools (a special class educated by a special education teacher within a regular school setting; Spencer, 2013), a form of less restrictive setting, was

lower compared to students in special schools, a more restrictive setting. Individual studies have also shown predominantly positive effects of inclusion on the cognitive outcomes of students with GLD (e.g., Gorges et al., 2018; Kocaj et al., 2014; Morvitz & Motta, 1992; Rea et al., 2002). The findings concerning psychosocial outcomes were more heterogeneous. For example, some studies showed that students with GLD in inclusive settings had higher self-concepts, greater feelings of competence, and less social avoidance and anxiety than students with GLD in segregated settings (e.g., Bakker et al., 2007; Peleg, 2011). Other studies showed that students with GLD in inclusive settings had lower self-concepts, had more anxiety, and felt less emotionally and socially integrated (e.g., Schmidt, 2000; Szumski & Karwowski, 2014).

Overall, students with GLD seem to benefit moderately from inclusive education. In particular, cognitive outcomes for students with GLD seem to be slightly more positive in more inclusive educational settings compared to settings that are more segregated.

### *Students Without GLD*

Several studies have summarized the effects of inclusive education for students with any kind of SEN for their peers without SEN (e.g., Kalambouka et al., 2007; Ruijs & Peetsma, 2009; Salend & Duhaney, 1999; Szumski et al., 2017). Most of these studies indicate neutral or slightly positive effects on cognitive and social outcomes among students without SEN in inclusive settings (Kalambouka et al., 2007; Salend & Duhaney, 1999). A recent meta-analysis investigated the effects of inclusive education on the academic achievement of students without SEN, finding an overall effect size of  $d = 0.12$  (Szumski et al., 2017). Subgroup analyses showed that the presence of students with mild SEN had slightly more positive effects on the performance of their peers without SEN ( $d = 0.19$ ) than the presence of students with severe SEN ( $d = 0.02$ ). None of these previous reviews reported specific effects of students with GLD on their peers without GLD.

In terms of cognitive outcomes, there is evidence that the inclusion of students with GLD typically has no or minimal effects on students without GLD (e.g., Bless & Klaghofer, 1991; Cole et al., 2004; Hienonen et al., 2018). In terms of psychosocial outcomes, studies have found no effects of including students with GLD on their peers without GLD (e.g., Arampatzi et al., 2011; Schwab, 2015). When students without GLD are taught together with students with GLD, there appear to be no detrimental effects on students without GLD.

### *The Present Meta-Analysis*

There is a research gap in the current empirical literature regarding the effects of inclusive education particularly for students with GLD as well as their peers without GLD. Some meta-analyses examining the effect of inclusive education on individual types of SEN and especially on students with GLD date back several decades (Carlberg & Kavale, 1980; Wang & Baker, 1985). A more recent meta-analysis examined the effects of inclusive education for students with all types of SEN together, without differentiating between individual types of SEN (Oh-Young & Filler, 2015). Moreover, there is currently only one meta-analysis on the effect



of inclusion on students without SEN (Szumski *et al.*, 2017). In this study, however, no detailed analysis of individual types of SEN was performed and only cognitive outcomes were considered. Hence, the present meta-analysis goes beyond existing single studies, reviews, and meta-analyses on the effects of inclusive education in several respects.

First, we specifically examined the effects of including students with GLD rather than students with all types of SEN. Second, we focused on outcomes for students with GLD as well as for their peers without GLD. Third, we examined both cognitive and psychosocial outcomes. Fourth, we investigated in two ways the possible selection effect that students with GLD with relatively high cognitive abilities are more likely to be educated in inclusive settings, while students with GLD with lower cognitive abilities are more likely to be taught in segregated settings. First, we examined whether the cognitive outcomes for students with GLD differ depending on study design (cross-sectional vs. longitudinal). While cross-sectional designs are susceptible to alternative explanations, longitudinal designs can make statements about the achievement development of students with GLD over time in different educational settings. More precisely, in cross-sectional studies, potential differences in outcomes between different educational settings might be due to initial selection more than the setting itself. In longitudinal studies, achievement development can be examined, allowing for clearer conclusions about placement effects. Second, we investigated differences in cognitive outcomes between studies that (a) employed matched samples; (b) controlled for factors such as IQ, age, and socioeconomic status; and (c) neither controlled for potential confounding variables nor matched samples of students with GLD in inclusive versus segregated settings. This was based on the following rationale: Matched samples assume that students with similar backgrounds (in terms of academic achievement, age, IQ, sex, socioeconomic status, associated difficulties) can be found in inclusive and segregated settings equally. In this case, there should be no selection effect. However, without prior control of the samples, there is the possibility of a selection effect: For example, better performing students might be more likely to be enrolled in inclusive settings than in segregated settings. A selection effect is also plausible with regard to psychosocial aspects, with more emotionally stable students with GLD more likely to be educated in inclusive educational settings, for example. However, an analysis of this latter point is not possible with the available data, because psychosocial aspects were not considered in the subgroups making up the matching moderator. Therefore, a possible selection effect is only investigated regarding cognitive outcomes among students with GLD.

We checked the robustness of the findings by considering further moderators regarding study-specific and sample characteristics. Study-specific characteristics were examined for cognitive as well as psychosocial outcomes among students with and without GLD. In order to examine possible changes in the implementation of inclusion over the years as well as between countries, we considered the publication year and the country where the studies were conducted as moderators. To investigate possible publication biases, we considered the publication status (unpublished, published) as a moderator in addition to the

usual approaches to detect bias. Sample characteristics that were examined concerning cognitive as well as psychosocial outcomes of students with and without GLD are age, school level, and diagnosis. Age and school level were considered as moderators in order to investigate a possible effect of students' age-related dependent and educational progress on the effects of inclusive education. Furthermore, as the criteria for diagnosing GLD differ between samples due to inconsistent criteria, we considered the clarity of GLD diagnosis as a moderator. Moreover, we divided the cognitive outcomes for students with and without GLD into subgroups. We examined whether the effect of inclusive education differed between mathematical and verbal achievement, as many studies differentiate between these outcomes (e.g., Cardona, 1997; Cole et al., 2004; Kocaj et al., 2014; Sharpe et al., 1994; Waldron & McLeskey, 1998). For example, it can be assumed that the verbal skills of students with GLD in inclusive education might be better than those of students with GLD in segregated settings because they have more verbal interactions with classmates without SEN and can thus implicitly improve their verbal skills. Furthermore, differences in expectations and curricula between inclusive and segregated educational settings are may be more or less pronounced in different subjects. With regard to psychosocial outcomes, a particularly large number of studies have examined students' self-concepts in more inclusive versus segregated educational settings (e.g., Cardona, 1997; Elbaum, 2002; Gorges et al., 2018; Sauer et al., 2007). In particular, given existing theoretical assumptions about self-concept (BIRG and BFLPE), we investigated the extent to which self-concept is influenced by inclusive education compared to other psychosocial factors.

In summary, we investigated the following questions via quantitative meta-analyses:

**Research Question 1:** Do students with GLD in inclusive education differ in cognitive aspects from students with GLD in segregated settings? (a) Can possible differences between the educational settings be attributed to a selection effect?

**Research Question 2:** Do students with GLD in inclusive education differ in psychosocial aspects from students with GLD in segregated settings?

**Research Question 3:** Do students without GLD in inclusive education, where they are taught together with students with GLD, differ in cognitive aspects from students without GLD in segregated settings where no students with GLD are present?

**Research Question 4:** Do students without GLD in inclusive education, where they are taught together with students with GLD, differ in psychosocial aspects from students without GLD in segregated settings where no students with GLD are present?

**Research Question 5:** Do the study-specific (publication year, country, publication status) or sample-specific moderators (age, school level, diagnosis, type of outcomes) influence the findings?

**Research Question 6:** Does a publication bias exist?

## Method

### *Information Retrieval*

Information retrieval took place via different methods: We searched several databases (EBSCOhost, ERIC, Sciencedirect, Ovid with PsychINFO and Psyn dex, FIS Bildung) and conducted a backward and forward search in relevant articles and previous reviews.

To systematically search these databases, we defined keywords that included synonyms for information about different educational settings and outcomes (e.g., [*“inclusion”* OR *“class placement”* OR *“special education”*]) AND [*“achievement”* OR *“effects”* OR *“self-concept”* OR *“social integration”*]; complete search term in supplemental material in the online version of the journal). We combined all educational setting terms with the outcomes and searched titles and abstracts in the databases. We conducted a not very restrictive search to ensure that as many relevant studies as possible were found. Furthermore, we screened the reference lists of relevant studies and previous reviews as a backward search and conducted a forward search by screening all studies that cited the relevant manuscripts.

### *Inclusion Criteria*

In order to investigate the influence of educational settings on cognitive outcomes and psychosocial aspects among students with GLD and their peers without GLD, we included studies that met the following inclusion criteria. A group of students must have been identified as having GLD. Due to the inconsistency of specific diagnostic criteria used across the primary studies, we included all studies explicitly examining students with GLD as the common ground. In some studies, GLD in several subjects were associated with general intellectual difficulty, defined as an average IQ between about 60 and 90 (e.g., Cardona, 1997; Dessemontet et al., 2012; Smogorzewska et al., 2019). Studies that did not report IQ were included when students were described as poorly achieving in more than one school subject (e.g., Bakker et al., 2007; Gorges et al., 2018; Koçaj et al., 2014; 2017; 2018; Szumski & Karwowski, 2014; 2015). Studies examining students with more severe general mental disabilities, with an IQ below 50, or specific learning difficulties in single subjects with an average IQ were excluded (e.g., Fore et al., 2008; Kennedy et al., 1997). In cases where it was still not entirely clear whether the sample met our criteria, a further code (unclear diagnostic criteria) was applied as a moderator variable (see section “Diagnosis” below). Studies were excluded in which different types of SEN were investigated collectively (examples of excluded studies because of irrelevant samples: Törmänen & Roebbers, 2018; Schwab et al., 2015).

We included only studies that compared students with and without GLD in more inclusive settings to a comparison group of students with and without GLD in segregated educational settings. From a policy perspective, inclusive education can be understood as the placement of students with disabilities together with students without disabilities in a joint educational setting. How this is implemented (e.g., extent of shared instruction, extent of special educational support) is often left open. Segregated education provides education for students with GLD separately from students without GLD, such as in special education schools.

Either cognitive (performance on standardized tests, metacognition) or psychosocial outcomes (social, attitudinal, emotional, and motivational aspects) must have been reported as dependent variables. In general, cognitive (learning) outcomes can be defined as “learning that is associated with knowledge of facts or processes” (IGI Global, n.d.). Cognitive learning outcomes are dependent on general cognitive skills such as working memory, problem solving ability, or meta-cognitive skills (Billing, 2007). These general cognitive skills are decisive for the development of domain-specific knowledge, which includes, for example, subject-related knowledge at school. This subject-related knowledge can be measured through, for example, multiple-choice, free-recall tasks, or free-sort tasks, which are often used in standardized achievement tests (e.g., Kraiger et al., 1993). Studies examining differences in cognitive outcomes between inclusive and segregated education mostly referred to subject-related knowledge. Therefore, studies that were included in the meta-analysis examined subject-related knowledge using standardized achievement tests, for example, in writing, reading comprehension, or mathematics (e.g., Cardona, 1997; Fruth & Woods, 2015; Waldron & McLeskey, 1998). One study included in the meta-analysis assessed the cognitive skill “metacognition” (Hessels & Schwab 2015). We did not consider school grades as a dependent variable because these depend on social frames of reference and are difficult to compare across classrooms (e.g., Brookhart, 1994; Brookhart et al., 2016).

Psychosocial is defined as “of or relating to the interrelation of social factors and individual thought and behaviour” (Oxford English Dictionary, n.d.), is mentioned as a noncognitive aspect (cf. Lipnevich & Roberts, 2014), and includes but is not limited to motivational, emotional, and attitudinal aspects (cf. Vasquez et al., 2016). We included studies examining psychosocial outcomes that could potentially change depending on the school setting. In order to narrow down the wide range of potential psychosocial outcomes, a literature search was conducted in advance to identify the psychosocial outcomes most commonly examined in potential primary studies investigating the effects of placement. Examples of included psychosocial outcomes were self-concept (e.g., Gorges et al., 2018; Sauer et al., 2007; Schwab, 2014), school leaving intention (Schwab, 2018), social integration (e.g., Bakker et al., 2007; Schmidt, 2000), aggressive or inappropriate behavior (e.g., Arampatzi et al., 2011; Martlew & Hodson, 1991), and well-being (e.g., Stelling, 2018). Personality traits were excluded because they are seen as relatively stable traits and considered to be less affected by school settings (e.g., Soto & Tackett, 2015; example of an excluded study: Porrata, 1997). In addition, studies that retrospectively measured outcomes such as social behavior were excluded to avoid bias (e.g., Klicpera & Gasteiger-Klicpera, 2006).

Included studies had to examine students in elementary or secondary school and had to be conducted between 1990 and 2019. Both content-related and methodological reasons justify the exclusion of studies before 1990. First, the ratification of the Convention on the Rights of the Child (United Nations, 1989) led segregated education to be increasingly seen as denying equal educational opportunities to students with disabilities (e.g., P. Alston et al., 1992). In addition, the American With Disabilities Act was passed in 1990 and emphasized avoiding discrimination and equal rights and opportunities for people with disabilities in

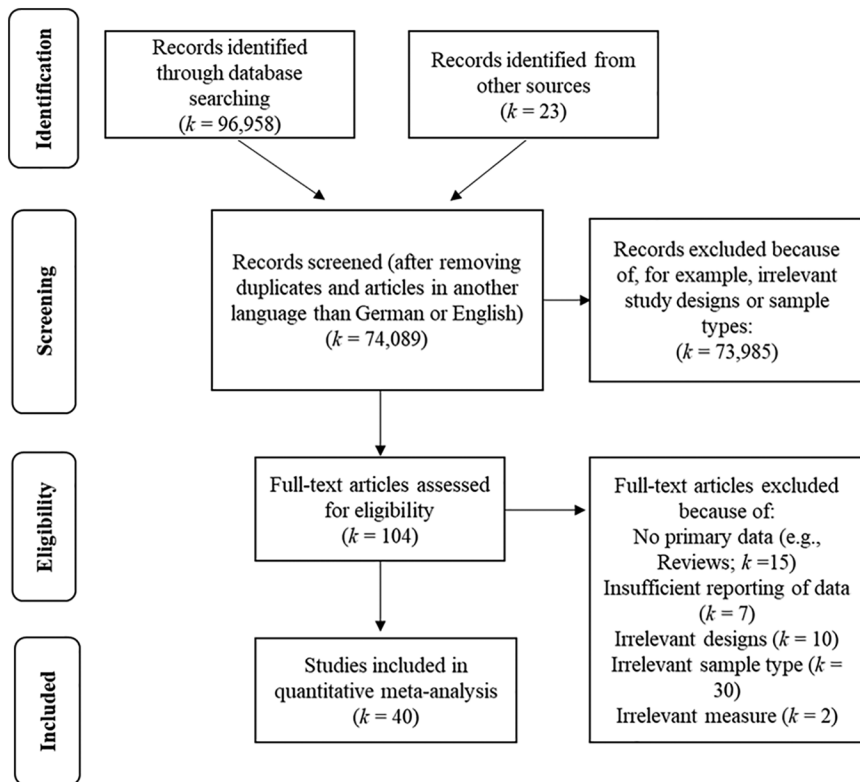


FIGURE 1. *Flow chart of the study selection process.*

the public sphere, including education and schooling. Thus, the early 1990s seem to have been a milestone in the implementation of inclusive education. Second, the full texts of studies conducted before 1990 were only rarely accessible despite all efforts made by the authors.

Only quantitative studies were included and the studies had to provide sufficient information to calculate effect sizes (examples of excluded studies due to lack of primary data availability: Peetsma et al., 2001; Ruijs et al., 2010). The full texts had to be available in English or German.

#### *Study Selection*

Figure 1 shows the study selection process and lists the total number of identified records from our information retrieval. In a first step, duplicates and articles in a language other than English or German were deleted, resulting in 74,089 studies. We then screened the titles and abstracts of all these records with respect to our inclusion criteria. We retrieved the full texts of studies that seemed to possibly fulfill our inclusion criteria. Some full-text articles were subsequently

excluded because closer assessment revealed that, for example, the sample was irrelevant or no primary data were reported. The full study selection procedure resulted in 40 studies matching all the inclusion criteria.

### *Data Coding*

We created a coding sheet with information about the included studies (e.g., country, study design), the samples (e.g., sample sizes in each group, identification of students with GLD), and the effect sizes (e.g., cognitive vs. psychosocial; longitudinal vs. cross-sectional). If no information concerning these aspects was found in the full text, it was coded as missing. All categories were defined precisely, and the variables were described using explicit criteria to ensure transparent coding between the two trained raters. Interrater reliability was calculated using Cohen's kappa for categorical variables (e.g., publication status, diagnosis, study design, outcome) and intra-class correlations for continuous variables (sample size, effect sizes). The two raters, who both coded all studies, exhibited an interrater agreement of Cohen's  $\kappa = 0.88$  with a range of  $0.70 < \kappa < 0.99$  for categorical variables. The mean intraclass correlation for continuous variables was .99. All differences in coding between the raters could be clarified by consulting the full texts.

### *Study-Specific Characteristics*

In order to control for study-specific characteristics, we coded the publication year, country where the study was conducted, publication status, as well as study design.

*Publication year.* We controlled for the publication year as a continuous moderator due to possible changes in the implementation of inclusive settings over the years, which might result in different effect sizes.

*Country.* Due to different country-specific guidelines for inclusive education, we coded the country in which the studies were conducted. The studies were conducted in many different countries, meaning that it was not possible to examine differences between individual countries. Since the implementation of inclusive education probably differs less within continents than between them, we summarized the countries into continents, specifically *North America* (coded as 0) and *Europe* (coded as 1). One study of students with GLD was conducted in Asia. However, since a single study from Asia is too small to be included as a moderator, we examined only North America and Europe. Since all studies examining psychosocial outcomes among students without GLD were conducted in Europe, no moderator analysis was calculated in this case.

*Publication status.* To check for possible publication bias, we coded the study type as either *unpublished* (0; e.g., dissertations) or *published work* (1; e.g., journal article, book chapter). Since all studies investigating cognitive outcomes among students with GLD and psychosocial outcomes of students without GLD were published, no moderator analyses were calculated in these cases.

*Design.* In some studies, the outcomes of inclusive education versus segregated education were measured at only one time point, while other studies reported longitudinal data. To check for difference in outcomes measured once compared to gains over time, we coded the study design as reporting either *cross-sectional* (0) or *longitudinal* (1) effects. Differences between longitudinal and cross-sectional effect sizes could also provide indications of a possible selection effect.

### *Sample Characteristics*

We assume that sample characteristics might influence effect sizes because the effectiveness of inclusive education might also depend on additional student characteristics. Therefore, we controlled for the students' age, school level, detailed description of GLD diagnosis, type of cognitive and psychosocial outcomes, and the extent to which the samples were matched.

*Age.* We considered the age of the students with GLD as a continuous variable, as it can be assumed that the age has an influence on the extent to which students benefit or do not benefit from inclusive education. Age for students without GLD was given only in two studies; thus, no moderator analyses were calculated here.

*School level.* We coded the school level as either *elementary school* (0) or *secondary school* (1) to check whether educational stage influences the effects of inclusive education. Elementary students ranged from age 6 to 11 and secondary students ranged from age 12 to 16. In none of the studies were students younger than 6 years or older than 16 years on average, which is why only elementary and secondary school students were included in the sample.

*Diagnosis.* The way students received their GLD diagnosis differed between samples. Some studies even did not provide criteria for the diagnosis. In particular, in some studies examining the effects of inclusion on students without GLD, the class composition regarding students with GLD was not specified clearly. Therefore, we controlled for the type of diagnosis as either diagnosis based on *transparent criteria* (0) or diagnosis based on *nontransparent criteria* (1).

*Type of cognitive outcome.* Inclusive education may have different effects on different types of cognitive outcomes. Therefore, we divided the cognitive outcomes into *mathematical performance* (0) and *verbal performance* (1).

*Type of psychosocial outcome.* Previous studies have investigated a variety of psychosocial variables that can be affected by students' educational setting. Self-concept has been particularly frequently examined and a number of theoretical assumptions relate to this, so we calculated a moderator analysis for *self-concept* (1) versus *other psychosocial outcomes* (0).

*Matching.* In order to check for a possible selection effect, we divided the primary studies into subgroups: whether the samples of students with GLD in inclusive schools versus students with GLD in segregated schools were fully matched

via *propensity score matching* (2), whether the samples were *controlled for cognitive abilities* (through IQ, age, gender, socio-economic status, associated difficulties, academic achievement; 1), or whether the samples were *neither matched nor controlled* (0).

### *Analyses*

During the coding process, we calculated Cohen's  $d$  for each outcome based on means, standard deviations, and sample sizes or the  $F$  values from analyses of variance reported in the primary studies. If students in more inclusive settings achieved more positive outcomes than students in segregated education, the effect size  $d$  was positive. A negative  $d$  resulted if students in segregated educational settings achieved outcomes that were more positive. Negatively coded psychosocial outcomes were recoded: For example, more aggression/anxiety in the segregated setting resulted in positive effect sizes.

Using R and the R packages *metafor* and *metaSEM*, we first calculated the variance of the effect sizes. Some primary studies were based on the same sample, so we summarized them in the analyses to control for dependencies (see Table 1). We calculated overall effect sizes for students with and without GLD separately. Then, we conducted four analyses: students with GLD–cognitive outcomes, students with GLD–psychosocial outcomes, students without GLD–cognitive outcomes, and students without GLD–psychosocial outcomes.

Due to the hierarchical data structure (various effect sizes within primary studies), we conducted a three-level meta-analysis (Cheung, 2015). More precisely, several dependent variables were investigated within each primary study. Standard errors for each effect size were estimated proportionally to the sample size. On the first level, the estimated standard errors of effect sizes served as sampling variance within the primary data (individual effect size level). Multiple effect sizes were reported in each primary study when multiple measures were evaluated for the same sample, for example, or when there were multiple measurement points. The second level thus concerned the variance in the effect sizes within each sample (effect sizes in studies). On the third level, effect sizes varied between primary studies (effect sizes between studies). Moderator analyses were also calculated separately for each of the four analyses using the three-level approach.

Furthermore, we carried out analyses to check for publication bias (cf. Pigott & Polanin, 2020; Polanin et al., 2016). Analyses of publication bias require average effect sizes per study, so we calculated mean effect sizes for each study and their variance proportional to sample size. We calculated classical random-effects models based on mean effect sizes. We then generated funnel plots to analyze the relationship between the effect sizes and their statistical power for each of the four subanalyses. We tested the significance of the funnel plots with Egger's test. In the case of a significant result in Egger's test, a  $p$  curve was examined (for details, see Simonsohn et al., 2014). Finally, we conducted sensitivity analyses to check the robustness of the effect sizes based on the three-level analyses.

## **Results**

Study selection resulted in  $k = 40$  studies with a total of 428 effect sizes and about  $N = 11,987$  examined students,  $n_1 = 6,119$  students with GLD and  $n_2 =$



**TABLE 1**  
*Overview of the studies included in the meta-analysis*

First author (year)	Country <sup>a</sup>	Design	Sample	DV	<i>N</i> <sub>inc</sub>	<i>N</i> <sub>seg</sub>	<i>d</i>	<i>V</i> ( <i>d</i> )	95% CI
1. Arampatzki (2011)	GR	CS	Peers	Psy	294–349	281–304	-0.10	.007	[-0.26, 0.06]
2. Bakker (2007)	NL	CS	GLD	Psy	74	213	0.24	.018	[-0.02, 0.50]
3. Bless (1991) <sup>1</sup>	CH	LT	Peers	Psy	175	175	0.09	.011	[-0.12, 0.30]
			Peers	Cog	175	175	0.10	.011	[-0.11, 0.31]
4. Brewton (2005)	US	CS	Peers	Cog	104	1002	-0.16	.011	[-0.36, 0.04]
5. Cardona (1997)	ES	LT	GLD	Psy	30	30	0.21	.067	[-0.29, 0.71]
			GLD	Cog	30	30	0.29	.067	[-0.22, 0.80]
6. Cole (2004)	US	LT	GLD	Cog	40–139	61–160	0.11	.020	[-0.17, 0.39]
7. Dessemontet (2012)	CH	LT	Peers	Cog	334	272	0.15	.007	[-0.01, 0.31]
			GLD	Psy	34	34	-0.21	.059	[-0.68, 0.26]
			GLD	Cog	34	34	0.20	.059	[-0.27, 0.67]
8. Fruth (2015)	US	CS	Peers	Cog	35–61	102–136	-0.25	.029	[-0.58, 0.08]
9. Gorges (2018) <sup>2</sup>	DE	LT	GLD	Psy	247	199	-0.00	.009	[-0.19, 0.19]
			GLD	Cog	247	199	0.66	.010	[0.47, 0.85]
10. Haeberlin (1990) <sup>1</sup>	CH	LT	GLD	Psy	9	9	-0.19	.035	[-0.56, 0.18]
			GLD	Cog	9	9	1.70	.121	[1.02, 2.38]
11. Hartfield (2009)	US	CS	Peers	Cog	192	270	-0.01	.009	[-0.19, 0.17]
12. Hessels (2015) <sup>3</sup>	AT	CS	GLD	Cog	32	11	0.19	.123	[-0.49, 0.87]
13. Hienonen (2018)	FI	CS	Peers	Cog	210	149	-0.68	.012	[-0.89, -0.47]
14. Kocaj (2014) <sup>4</sup>	DE	CS	GLD	Cog	278–282	241–258	0.42	.008	[0.25, 0.59]
15. Kocaj (2017)	DE	CS	GLD	Cog	784–793	486–488	0.59	.003	[0.48, 0.70]
16. Kocaj (2018) <sup>4</sup>	DE	CS	GLD	Psy	289	261	-0.45	.007	[-0.62, -0.28]

(continued)

**TABLE 1 (continued)**

First author (year)	Country <sup>a</sup>	Design	Sample	DV	$N_{inc}$	$N_{seg}$	$d$	$V(d)$	95% CI
17. Marston (1996)	US	LT	GLD	Cog	33–36	36–171	0.20	.043	[-0.20, 0.60]
18. Martlew (1991)	NO	CS	GLD	Psy	10	18	-0.01	.156	[-0.78, 0.76]
19. Morvitz (1992)	US	CS	GLD	Psy	35	31	0.18	.061	[-0.30, 0.66]
20. Murawski (2006)	US	LT	GLD	Cog	35	31	1.07	.070	[0.56, 1.58]
			Peers	Cog	26	16	-0.16	.101	[-0.78, 0.46]
			GLD	Cog	4–8	14	0.22	.239	[-0.73, 1.17]
21. Nusser (2016)	DE	CS	GLD	Psy	128–131	466–469	-0.19	.010	[-0.38, 0.00]
22. Peleg (2011)	IL	CS	GLD	Psy	24	16	0.98	.116	[0.32, 1.64]
23. Pescatore (2000)	US	CS	GLD	Psy	9	8	0.35	.240	[-0.60, 1.30]
24. Porlier (1999)	CA	CS	GLD	Psy	115	112	-0.04	.018	[-0.30, 0.22]
25. Rathmann (2018)	DE	CS	GLD	Psy	407	126	0.05	.010	[-0.15, 0.25]
26. Rea (2002)	US	LT	GLD	Psy	35	21	0.67	.080	[0.12, 1.22]
			GLD	Cog	34–35	21	0.29	.078	[-0.25, 0.83]
27. Rossmann (2011)	AT	CS	GLD	Psy	52	56	0.12	.037	[-0.26, 0.50]
28. Sauer (2007)	DE	CS	GLD	Psy	154	516	-0.13	.008	[-0.31, 0.05]
29. Schmidt (2000)	SI	CS	GLD	Psy	19	21	-1.03	.114	[-1.69, -0.37]
30. Schwab (2014) <sup>3</sup>	AT	CS	GLD	Psy	32	11	0.43	.124	[-0.26, 1.12]
31. Schwab (2015)	AT	CS	Peers	Psy	485	501	0.08	.004	[-0.04, 0.20]
32. Schwab (2018)	AT	CS	Peers	Psy	564	483	0.13	.004	[0.01, 0.25]
33. Sharpe et al. (1994)	US	LT	Peers	Cog	35	108	-0.11	.038	[-0.49, 0.27]

(continued)

TABLE 1 (continued)

First author (year)	Country <sup>a</sup>	Design	Sample	DV	$N_{inc}$	$N_{seg}$	$d$	$V(d)$	95% CI
34. Smogorzewska (2019)	PL	LT	GLD	Psy	79	87	0.05	.024	[-0.25, 0.35]
			GLD	Cog	79	87	0.15	.024	[-0.15, 0.45]
35. Stelling (2018)	DE	CS	GLD	Psy	74	75	0.09	.027	[-0.23, 0.41]
36. Stranghøner (2017) <sup>2</sup>	DE	LT	GLD	Cog	227–239	158–171	0.52	.011	[0.32, 0.72]
37. Szumski (2014)	PL	CS	GLD	Psy	256–294	309	-0.26	.007	[-0.42, -0.10]
			GLD	Cog	256–294	309	0.22	.007	[0.06, 0.38]
38. Szumski (2015)	PL	CS	GLD	Psy	410	195	-0.37	.008	[-0.54, -0.20]
			GLD	Cog	410	195	0.41	.008	[0.24, 0.58]
39. Waldron (1998)	US	LT	GLD	Cog	71	73	0.06	.028	[-0.27, 0.39]
40. Wild (2015) <sup>2</sup>	DE	CS	GLD	Psy	189	181	-0.01	.011	[-0.21, 0.19]
			GLD	Cog	189	181	0.99	.012	[0.78, 1.20]

Note. <sup>1,2,3,4</sup>Studies were summarized in the analyses because they used the same sample.  $d$  = mean effect size of the studies in terms of Cohen's  $d$ . GLD = general learning difficulties; DV = dependent variable;  $N_{inc}$  = sample size in the inclusive setting;  $N_{seg}$  = sample size in the segregated settings; CS = cross-sectional; LT = longitudinal, Psy = psychosocial variables; Cog = cognitive variables; CI = confidence interval.

<sup>a</sup>Country codes according to ISO 3166 ALPHA-2.

5,868 students without GLD. Table 1 gives an overview of all studies, including the country in which the studies were conducted, the design (cross-sectional or longitudinal), the sample (students with and/or without GLD), and the type of outcome measured (cognitive, psychosocial). Furthermore, it presents the sample sizes within the *inclusive education* and *segregated education* groups as well as average effect sizes, their variance, and the 95% confidence intervals (CIs).

### *Effect Sizes From Overall Analyses*

We first analyzed the overall effect sizes in a three-level approach. Within the subgroup of students with GLD, students in inclusive settings outperformed their counterparts in segregated settings ( $d = 0.14$ ,  $SE = 0.05$ , 95% CI [0.05, 0.24], range =  $-1.77 < d < 2.36$ ,  $p < .01$ ). More important, the effect for cognitive outcomes among students with GLD (Research Question 1) was positive and statistically significant ( $d = 0.35$ ,  $SE = 0.09$ , CI [0.18, 0.52], range =  $-1.77 < d < 2.36$ ,  $p < .001$ ), while the effect for psychosocial outcomes (Research Question 2) was not statistically significant ( $d = 0.00$ ,  $SE = 0.07$ , CI [-0.14, 0.15], range =  $-1.39 < d < 1.33$ ,  $p = .95$ ). With regard to cognitive outcomes, students with GLD in inclusive education outperformed their counterparts in segregated settings. With regard to psychosocial outcomes, we found no differences between the two settings for students with GLD.

For students without GLD, we found no significant effects ( $d = -0.08$ ,  $SE = 0.07$ , 95% CI [-0.21, 0.05], range =  $-1.07 < d < 1.10$ ,  $p = .24$ ) for either cognitive outcomes (Research Question 3;  $d = -0.14$ ,  $SE = 0.09$ , 95% CI [-0.32, 0.04], range =  $-1.07 < d < 1.10$ ,  $p = .14$ ) or for psychosocial outcomes (Research Question 4;  $d = 0.06$ ,  $SE = 0.05$ , CI [-0.04, 0.16], range =  $-0.35 < d < 0.45$ ,  $p = .22$ ).

The heterogeneity of variance not attributable to sampling error was high ( $73.07\% < I^2 < 90.61\%$ ) in each subanalysis (see Tables 2–5). For cognitive outcomes among students with GLD, the heterogeneity of effect sizes between different measures within primary studies was relatively low ( $I^2$  on Level 2) and between primary studies relatively moderate ( $I^2$  on Level 3). For psychosocial outcomes among students with GLD, the heterogeneity between different measures within primary studies was moderate and the heterogeneity of effect sizes between primary studies was low to moderate. For students without GLD, the heterogeneity of effect sizes not attributable to sampling error for cognitive outcomes was moderate between measures within primary studies and relatively low between primary studies. For psychosocial outcomes among students without GLD, there was no heterogeneity between different measures within primary studies beyond the heterogeneity due to sampling error and a very high heterogeneity between primary studies. The relatively large variances indicate that moderators must be considered.

### *Moderator Analyses*

We included four study-specific (publication year, country, publication status, study design) and five sample-specific moderators (age, school level, diagnosis, type of outcome, matching) regarding Research Question 5. Table 2 shows the results for cognitive outcomes and Table 3 for psychosocial outcomes, both

**TABLE 2**

*Results of the overall model as well as moderator analyses for the effects on cognitive outcomes of students with GLD*

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Overall model	16	0.35	***	.09	[0.18, 0.52]
$I^2(2)/I^2(3)$				25.45/65.17	
$\tau^2(2)/\tau^2(3)$				0.07/0.16	
Moderators					
Publication year <sup>a</sup>	16				
Intercept		0.40	***	.10	[0.20, 0.60]
Year (difference)		-0.08		.09	[-0.25, 0.09]
$I^2(2)/I^2(3)$				30.12/61.03	
$\tau^2(2)/\tau^2(3)$				0.09/0.18	
Age <sup>a</sup>	11				
Intercept		0.32	***	.06	[0.19, 0.44]
Age (difference)		-0.04		.06	[-0.16, 0.07]
$I^2(2)/I^2(3)$				2.56/83.24	
$\tau^2(2)/\tau^2(3)$				0.01/0.22	
Country					
North America (RC)	6	0.19		.12	[-0.05, 0.43]
Europe	10	0.45	***	.16	[0.25, 0.64]
Subgroup difference		-0.26			
$I^2(2)/I^2(3)$				18.98/70.90	
$\tau^2(2)/\tau^2(3)$				0.05/0.19	
Design <sup>b</sup>					
Cross-sectional (RC)	16	<b>0.48</b>	***	.09	[0.30, 0.66]
Longitudinal	9	<b>0.05</b>		.10	[-0.16, 0.25]
Subgroup difference		<b>0.44</b>	***		
$I^2(2)/I^2(3)$				34.24/55.17	
$\tau^2(2)/\tau^2(3)$				0.09/0.14	
School level					
Elementary (RC)	10	0.40	**	.13	[0.14, 0.65]
Secondary	3	0.38		.25	[-0.12, 0.88]
Subgroup difference		0.02			
$I^2(2)/I^2(3)$				36.19/56.40	
$\tau^2(2)/\tau^2(3)$				0.12/0.19	
Diagnosis					
Diagnosis clear (RC)	12	0.42	***	.09	[0.25, 0.59]
Diagnosis unclear	4	0.19		.12	[-0.04, 0.41]
Subgroup difference		0.23			
$I^2(2)/I^2(3)$				16.98/72.66	
$\tau^2(2)/\tau^2(3)$				0.04/0.19	

*(continued)*

**TABLE 2 (continued)**

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Type of outcome					
Mathematical (RC)	9	0.32	**	.10	[0.13, 0.51]
Verbal	11	0.32	**	.10	[0.13, 0.51]
Subgroup difference					
$I^2(2)/I^2(3)$		0.00			
$\tau^2(2)/\tau^2(3)$				26.35/64.41	
				0.08/0.19	
Matching					
No matching	5	0.29	*	.14	[0.02, 0.57]
Controlled	7	0.21		.13	[-0.05, 0.47]
Full matching	4	0.64	***	.16	[0.32, 0.96]
$I^2(2)/I^2(3)$				23.15/67.14	
$\tau^2(2)/\tau^2(3)$				0.06/0.19	

Note. Values in bold represent significant subgroup differences.  $I^2(2) = I^2$  for Level 2,  $I^2(3) = I^2$  for Level 3. GLD = general learning difficulties; RC = reference category; *k* = number of samples; Sig = significance of the *t* values; CI = confidence interval.

<sup>a</sup>Continuous and centered. <sup>b</sup>Analyses at effect size level instead of study level: Longitudinal studies include both cross-sectional and longitudinal effect sizes.

\* $p < .05$ . \*\* $p \leq .01$ . \*\*\* $p \leq .001$ .

**TABLE 3**

Results of the overall model as well as moderator analyses for the effects on psychosocial outcomes of students with GLD

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Overall model	22	0.00		.07	[-0.14, 0.15]
$I^2(2)/I^2(3)$				52.09/32.39	
$\tau^2(2)/\tau^2(3)$				0.09/0.06	
Moderators					
Publication year <sup>a</sup>	22				
Intercept		0.01		.09	[-0.16, 0.19]
Year (difference)		-0.01		.07	[-0.14, 0.12]
$I^2(2)/I^2(3)$				53.95/31.12	
$\tau^2(2)/\tau^2(3)$				0.10/0.06	
Age <sup>a</sup>	16				
Intercept		0.03		.08	[-0.13, 0.20]
Age (difference)		0.12		.09	[-0.07, 0.30]
$I^2(2)/I^2(3)$				39.08/41.38	
$\tau^2(2)/\tau^2(3)$				0.06/0.08	
Country					
North America (RC)	4	0.20		.16	[-0.13, 0.52]
Europe	17	-0.08		.07	[-0.21, 0.06]

(continued)

**TABLE 3 (continued)**

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Subgroup difference		0.27			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				39.19/41.37	
$\tau^2$ (2)/ $\tau^2$ (3)				0.06/0.06	
Publication status					
Unpublished (RC)	2	0.20		.24	[-0.28, 0.68]
Published	20	-0.02		.08	[-0.17, 0.14]
Subgroup difference		0.21			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				52.89/31.84	
$\tau^2$ (2)/ $\tau^2$ (3)				0.10/0.06	
Design <sup>b</sup>					
Cross-sectional (RC)	21	0.02		.08	[-0.13, 0.17]
Longitudinal	5	-0.07		.10	[-0.26, 0.13]
Subgroup difference		0.08			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				52.93/31.74	
$\tau^2$ (2)/ $\tau^2$ (3)				0.10/0.06	
School level					
Elementary (RC)	9	-0.14		.12	[-0.38, 0.10]
Secondary	12	0.17		.10	[-0.04, 0.37]
Subgroup difference		-0.30			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				52.62/31.62	
$\tau^2$ (2)/ $\tau^2$ (3)				0.09/0.06	
Diagnosis					
Diagnosis clear (RC)	17	0.02		.09	[-0.15, 0.19]
Diagnosis unclear	5	-0.05		.17	[-0.39, 0.28]
Subgroup difference		0.07			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				53.70/31.30	
$\tau^2$ (2)/ $\tau^2$ (3)				0.10/0.06	
Type of outcome					
Self-concept (RC)	10	0.08		.09	[-0.09, 0.25]
Others	17	-0.03		.07	[-0.19, 0.12]
Subgroup difference		0.11			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)				53.06/31.72	
$\tau^2$ (2)/ $\tau^2$ (3)				0.10/0.06	

*Note.* Values in bold represent significant subgroup differences. *I*<sup>2</sup>(2) = *I*<sup>2</sup> for Level 2, *I*<sup>2</sup>(3) = *I*<sup>2</sup> for Level 3. GLD = general learning difficulties; RC = reference category, *k* = number of samples, Sig = significance of the *t* values; CI = confidence interval.

<sup>a</sup>Continuous and centered. <sup>b</sup>Analyses at effect size level instead of study level: Longitudinal studies include both cross-sectional and longitudinal effect sizes.

**TABLE 4**

*Results of the overall model as well as moderator analyses for the effects on cognitive outcomes for students without GLD*

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Overall model	8	-0.14		.09	[-0.32, 0.04]
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			49.03/24.04		
$\tau^2$ (2)/ $\tau^2$ (3)			0.05/0.02		
Moderators					
Publication year <sup>a</sup>	8				
Intercept		<b>-0.10</b>		.08	[-0.25, 0.06]
Year		<b>-0.15</b>	*	.07	[-0.29, -0.01]
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			37.16/29.29		
$\tau^2$ (2)/ $\tau^2$ (3)			0.03/0.02		
Country					
North America (RC)	6	-0.09		.10	[-0.29, 0.12]
Europe	2	-0.29		.19	[-0.66, 0.08]
Subgroup differences		0.20			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			49.37/23.87		
$\tau^2$ (2)/ $\tau^2$ (3)			0.05/0.02		
Publication status					
Unpublished (RC)	2	-0.08		.22	[-0.52, 0.35]
Published	6	-0.15		.11	[-0.36, 0.07]
Subgroup differences		0.06			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			52.88/22.31		
$\tau^2$ (2)/ $\tau^2$ (3)			0.06/0.02		
Design					
Cross-sectional (RC)	8	<b>-0.18</b>	*	.08	[-0.35, -0.02]
Longitudinal	4	<b>0.06</b>		.10	[-0.14, 0.26]
Subgroup differences		<b>-0.24</b>	**		
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			46.70/21.18		
$\tau^2$ (2)/ $\tau^2$ (3)			0.04/0.02		
School level					
Elementary (RC)	2	<b>0.04</b>		.11	[-0.18, 0.26]
Secondary	5	<b>-0.26</b>	**	.10	[-0.45, -0.07]
Subgroup differences		<b>0.30</b>	*		
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			34.95/30.97		
$\tau^2$ (2)/ $\tau^2$ (3)			0.03/0.03		
Diagnosis					
Diagnosis clear (RC)	3	-0.03		.13	[-0.29, 0.23]
Diagnosis unclear	6	-0.18		.10	[-0.38, 0.02]
Subgroup differences		0.15			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			52.10/22.28		
$\tau^2$ (2)/ $\tau^2$ (3)			0.06/0.02		

(continued)



**TABLE 4 (continued)**

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Type of outcome					
Mathematical	6	-0.11		.10	[-0.30, 0.08]
Verbal	4	-0.11		.10	[-0.30, 0.08]
Subgroup difference		0.00			
<i>F</i> <sup>2</sup> (2)/ <i>F</i> <sup>2</sup> (3)					46.26/25.92
$\tau^2$ (2)/ $\tau^2$ (3)					0.05/0.03

Note. Values in bold represent significant subgroup differences. *F*<sup>2</sup>(2) = *F*<sup>2</sup> for Level 2, *F*<sup>2</sup>(3) = *F*<sup>2</sup> for Level 3. GLD = general learning difficulties; RC = reference category; *k* = number of samples; sig = significance of the *t* values; CI = confidence interval.

<sup>a</sup>Continuous and centered. <sup>b</sup>Analyses at effect size level instead of study level: longitudinal studies include both cross-sectional and longitudinal effect sizes.

\**p* < .05. \*\**p* ≤ .01. \*\*\**p* ≤ .001.

**TABLE 5**

Results of the overall model as well as moderator analyses for the effects on psychosocial outcomes for students without GLD

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Overall model	4	0.06		.05	[-0.04, 0.16]
<i>F</i> <sup>2</sup> (2)/ <i>F</i> <sup>2</sup> (3)					0.00/83.27
$\tau^2$ (2)/ $\tau^2$ (3)					0.00/0.04
Moderators					
Publication year <sup>a</sup>	4				
Intercept		0.06		.05	[-0.04, 0.16]
Year		-0.01		.03	[-0.08, 0.06]
<i>F</i> <sup>2</sup> (2)/ <i>F</i> <sup>2</sup> (3)					0.00/83.99
$\tau^2$ (2)/ $\tau^2$ (3)					0.00/0.04
Design					
Cross-sectional (RC)	4	0.07		.05	[-0.04, 0.18]
Longitudinal	1	0.02		.11	[-0.22, 0.26]
Subgroup differences		0.05			
<i>F</i> <sup>2</sup> (2)/ <i>F</i> <sup>2</sup> (3)					0.00/83.91
$\tau^2$ (2)/ $\tau^2$ (3)					0.00/0.04
School level					
Elementary (RC)	2	0.06		.07	[-0.08, 0.20]
Secondary	3	0.06		.08	[-0.10, 0.22]
Subgroup differences		0.00			
<i>F</i> <sup>2</sup> (2)/ <i>F</i> <sup>2</sup> (3)					0.86/83.30
$\tau^2$ (2)/ $\tau^2$ (3)					0.00/0.04
Diagnosis					
Diagnosis clear (RC)	3	0.03		.07	[-0.12, 0.18]
Diagnosis unclear	1	0.08		.07	[-0.05, 0.22]

(continued)

**TABLE 5 (continued)**

Models	<i>k</i>	<i>d</i>	Sig	<i>SE</i>	95% CI
Subgroup differences		-0.05			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			0.00/83.85		
$\tau^2$ (2)/ $\tau^2$ (3)			0.00/0.04		
Type of outcome					
Self-concept	1	0.05		.14	[-0.25, 0.35]
Others	5	0.06		.05	[-0.05, 0.17]
Subgroup difference		-0.02			
<i>I</i> <sup>2</sup> (2)/ <i>I</i> <sup>2</sup> (3)			0.00/84.01		
$\tau^2$ (2)/ $\tau^2$ (3)			0.00/0.04		

*Note.* Values in bold represent significant subgroup differences. *I*<sup>2</sup>(2) = *I*<sup>2</sup> for Level 2, *I*<sup>2</sup>(3) = *I*<sup>2</sup> for Level 3. GLD = general learning difficulties; RC = reference category; *k* = number of samples; Sig = significance of the *t* values; CI = confidence interval.

<sup>a</sup>Continuous and centered. <sup>b</sup>Analyses at effect size level instead of study level: Longitudinal studies include both cross-sectional and longitudinal effect sizes.

among students with GLD. Concerning cognitive outcomes (Research Question 1a), study design had a moderating effect: The effect sizes in cross-sectional studies were higher than the slightly positive nonsignificant effects in longitudinal studies,  $F(1, 159) = 30.68, p < .001$ . There were no significant differences in the effect sizes between the three groups of studies employing propensity score matching versus controlling for cognitive abilities versus no matching,  $F(2, 179) = 2.18, p = .12$ . The other moderators did not yield significant effects (see Table 2). For psychosocial outcomes, no moderators were significant (Table 3).

Table 4 shows the results for cognitive outcomes and Table 5 for psychosocial outcomes among students without GLD. For cognitive outcomes, publication year had a moderating effect: The older the study, the higher the effect sizes,  $F(1, 56) = 4.45, p = .04$ . Furthermore, study design was a moderator, with a significantly negative effect size in cross-sectional studies and a slightly positive nonsignificant effect size in longitudinal studies,  $F(1, 56) = 10.44, p < .01$ . School level was the third significant moderator, with a significant negative effect size at the secondary school level and a nonsignificant effect size at the elementary school level,  $F(1, 56) = 4.27, p = .04$  (see Table 4). No moderators were found concerning psychosocial outcomes among students without GLD (Table 5).

#### *Publication Bias*

To answer Research Question 6, we tested for the presence of publication bias in three ways. First, we considered publication status as a possible moderator in the analyses. We found no significant influence of publication status on the level of effect sizes (see Tables 2–5).

Second, we conducted funnel plots for each subanalysis (Figures 2–5), inspected them visually, and calculated Egger's tests to determine significance. Egger's test showed no significant asymmetry for studies on cognitive outcomes among students with GLD ( $z = 0.27, p = .79$ ) and students without GLD ( $z = -0.39, p = .70$ ) as well as studies on psychosocial outcomes among

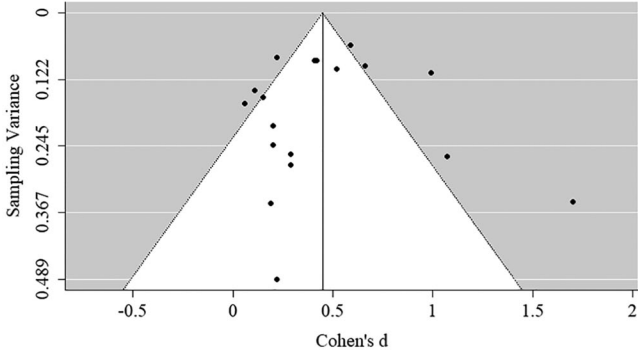


FIGURE 2. *Funnel plot for cognitive outcomes for students with general learning difficulties.*

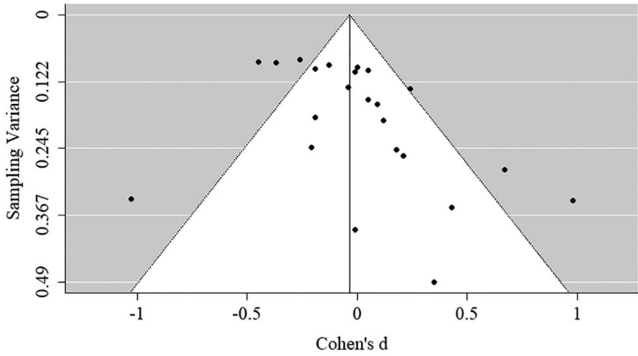


FIGURE 3. *Funnel plot for psychosocial outcomes for students with general learning difficulties.*

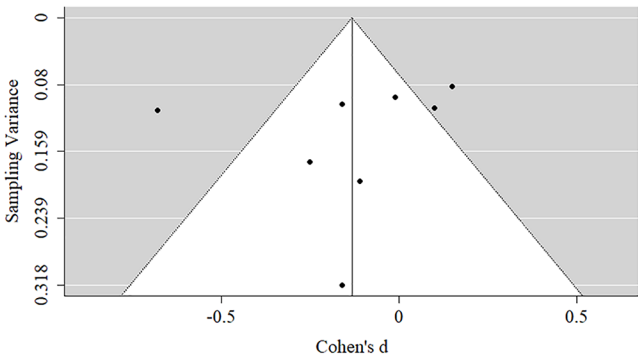


FIGURE 4. *Funnel plot for cognitive outcomes for students without general learning difficulties.*

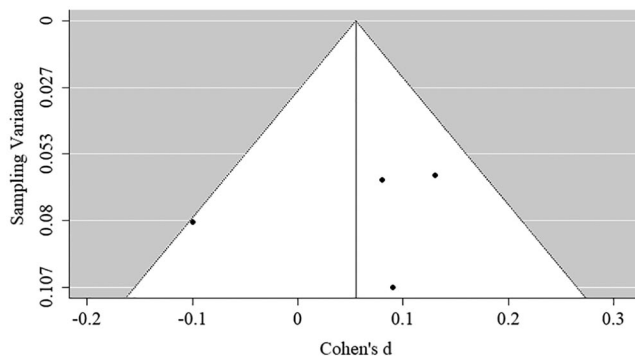


FIGURE 5. *Funnel plot for psychosocial outcomes for students without general learning difficulties.*

students without GLD ( $z = -0.42, p = .68$ ). However, for studies on psychosocial outcomes among students with GLD, a significant asymmetry was shown ( $z = 2.50, p = .01$ ). The results of Egger's test were in line with visual inspection concerning asymmetry. Further analyzing the effect sizes with significant  $p$  values, a significantly right-skewed  $p$  curve was found (half  $p$  curve,  $z = -5.35, p < .001$ ; full  $p$  curve,  $z = -4.38, p < .001$ ): There were more high than low significant  $p$  values (for details on  $p$  curves, see Simonsohn et al., 2014).

Third, we conducted sensitivity analyses to examine the robustness of the effect sizes (see Table 6). To do so, we excluded studies with the highest mean effect sizes. For cognitive outcomes among students with GLD, we excluded the effect sizes from Haeberlin et al. (1990) and then from both Haeberlin et al. (1990) and Morvitz and Motta (1992). For psychosocial outcomes among students with GLD, we first excluded Schmidt (2000) and then also Peleg (2011). Due to the low number of studies concerning students without GLD, we only excluded one study for each outcome with the highest mean effect size: Hienonen et al. (2018) for cognitive outcomes and Schwab (2018) for psychosocial outcomes. The sensitivity analyses revealed no changes in the significance of the effect sizes, indicating highly robust effect sizes.

## Discussion

The present meta-analysis examined the effects of inclusive education versus segregated education on cognitive and psychosocial outcomes for students with and without GLD. In this section, we first discuss the findings on cognitive outcomes, including selection effects, and then on psychosocial outcomes among students with GLD. We then proceed to discuss the findings for students without GLD, first for cognitive outcomes, then for psychosocial outcomes. Subsequently, we discuss implications for practice. Finally, we highlight limitations and implications for future research and end with a conclusion.

**TABLE 6**  
*Robustness check of effect sizes with outliers excluded*

Models	Students with GLD						Students without GLD									
	Cognitive			Psychosocial			Cognitive			Psychosocial						
	<i>k</i>	<i>d</i>	SE	95% CI	<i>k</i>	<i>d</i>	SE	95% CI	<i>k</i>	<i>d</i>	SE	95% CI				
Overall effect	16	0.35***	0.09	[0.18, 0.52]	22	0.00	0.07	[-0.14, 0.15]	8	-0.14	0.09	[-0.32, 0.04]	4	0.06	0.05	[-0.04, 0.16]
One outlier excluded	15	0.28***	0.06	[0.17, 0.40]	21	0.06	0.07	[-0.07, 0.20]	7	-0.10	0.07	[-0.24, 0.05]	3	0.04	0.05	[-0.06, 0.15]
Two outliers excluded	14	0.28***	0.06	[0.17, 0.39]	20	0.03	0.05	[-0.08, 0.13]								

*Note.* Studies with the same sample were summarized in these analyses. GLD = general learning difficulties; CI = confidence interval.  
 \**p* < .05. \*\**p* ≤ .01. \*\*\**p* ≤ .001.

### *Students With GLD*

Regarding Research Question 1, the results showed that students with GLD benefited from more inclusive education in terms of cognitive outcomes. They exhibited better academic performance compared to their counterparts in more segregated settings. We extended existing research by focusing explicitly on students with GLD and considering more recent research. Our results regarding cognitive outcomes for students with GLD are in line with previous meta-analyses: Carlberg and Kavale (1980) as well as Wang and Baker (1985) also found positive effects on cognitive outcomes for inclusive education compared to segregated education among students with GLD and intellectual disabilities, respectively. A meta-analysis of students with all types of SEN showed comparable results (Oh-Young & Filler, 2015). Moreover, the positive effect seems to be consistent across different cognitive outcomes. First, taking the different levels of the meta-analysis into account, the effect sizes for cognitive outcomes are largely consistent across different measures within primary studies (Level 2). Many primary studies examine cognitive outcomes in various subjects, such as mathematical and verbal outcomes. Second, a moderator analysis also showed no difference in effect sizes between mathematical and verbal outcomes. This is in line with previous research showing correlations between different achievement outcomes (for an overview regarding mathematics and language outcomes, see Peng et al., 2020). In summary, the finding that students with GLD cognitively benefit from inclusive education seems to be consistent both across different cognitive outcomes and across previous meta-analyses.

These findings support the United Nations' (2006) decision to call for an inclusive school system. Inclusive classes appear to have an advantage over segregated classes because they provide, for example, a more stimulating learning environment and place greater emphasis on students' performance (e.g., Cole et al., 2004; Myklebust, 2007). Furthermore, composition effects seem to emerge: Students with GLD in higher performing classes (e.g., inclusive settings) perform better than students with GLD in generally lower performing classes (e.g., segregated classes for students with GLD; e.g., Harker & Tymms, 2004).

However, before concluding that inclusive education itself leads to higher cognitive outcomes among students with GLD, we additionally examined the presence of a possible selection effect (Research Question 1a). First, we found evidence that cognitive outcomes for students with GLD were moderated by the study design. In cross-sectional studies, students with GLD in inclusive settings outperformed students with GLD in segregated educational settings. However, increases in performance over time (longitudinally) were similar in inclusive and segregated settings. Once students with GLD are placed in a given educational setting, their individual learning gains do not seem to differ (even if students with GLD perform better in inclusive schools overall). Selection effects in which the choice of school depends on students' previous performance could explain these findings. More precisely, it can be assumed that students with GLD with comparatively high abilities are more likely to be enrolled in inclusive education, while their counterparts with lower academic abilities are more likely to be enrolled in segregated schools (e.g., Dessemontet et al., 2012; Madden & Slavin, 1983;

Möller, 2013). A reason for this could be that better performing students with GLD are better able to meet the higher requirements at inclusive schools (cf. Madden & Slavin, 1983). Moreover, in order to correctly interpret the findings, the time that students had already spent in inclusive education at the cross-sectional studies' single measurement point must be taken into account. Perhaps the students had been in their respective educational settings for so long by the time their performance was measured that this already reflected a performance improvement resulting from the setting. Performance would have to be tested directly before students' transition to inclusive versus segregated educational settings in order to make a clear statement about selection effects. Since most studies did not clearly indicate the time interval between enrollment in each educational setting and the measurement point, our ability to make conclusions on selection effects is limited. Thus, in order to be able to make more precise statements about selection effects, we then examined the extent to which each study employed sample matched. Whether a sample employed full propensity score matching, controlled for cognitive abilities, or involved no matching or control had no significant effect on the outcome variables. This finding speaks against a selection effect, although it must be noted that only a few studies had matched samples, resulting in a high standard error. Selection effects cannot be ruled out in general, as students are not randomly distributed into different educational settings in practice. While the selection of educational setting is the responsibility of different actors, and in most countries, children without SEN are usually assigned to a school by the local school district (e.g., Jacobs, 2011), parents also have a say and can influence this decision by, for example, taking into account information provided by special education and other teachers (e.g., Pyryt & Bosetti, 2007). Moreover, there are indications that factors such as students' social background can influence school decisions. Previous studies have shown that attending a regular school depends on social background (e.g., Kocaj *et al.*, 2014; Kölm *et al.*, 2017; Szumski & Karwowski, 2012) as well as achievement-related aspects (e.g., Bless & Mohr, 2007; Dessemontet *et al.*, 2012). Students with SEN with higher socioeconomic status and with comparatively better academic achievement as well as higher IQs seem to be more likely to attend regular schools, while students with SEN with lower socioeconomic status and worse academic achievement as well as a lower IQ more often attend segregated educational settings. The extent to which factors such as previous performance might influence the choice of school in the sense of a selection effect and which decision-makers influence these aspects remains an open question and should be the subject of future research.

Regarding Research Question 2, the finding that psychosocial outcomes for students with GLD do not differ in inclusive settings compared to segregated settings does not speak against the implementation of inclusive education. Students with GLD in more inclusive settings did not suffer in terms of self-concept or emotional aspects compared with their peers in segregated settings. Previous research results regarding the psychosocial effects of inclusive education were quite heterogeneous. For example, Carlberg and Kavale (1980) showed more positive psychosocial outcomes in inclusive education for students with GLD.

Psychosocial outcomes in the examined primary studies included not only self-concept and social acceptance but also personality and behavior. Wang and Baker (1985) detected no differences between inclusive and segregated education when including primary studies considering mainly attitudinal outcomes and self-concept. One reason for these different results could be the high heterogeneity in psychosocial outcomes available for study. Different factors were included in the previous meta-analyses, and the effects of individual outcomes may have canceled each other out. This assumption is supported by the rather moderate variation in effect sizes between different measures within primary studies (Level 2) as well as between primary studies (at Level 3 in our three-level meta-analysis), which implies that primary studies investigated different psychosocial aspects. In order to make more precise statements about the effect of inclusion on individual psychosocial factors, however, many more studies are needed.

Furthermore, it must be taken into account that in our analyses, a variety of psychosocial outcomes were considered together. It is possible that individual psychosocial outcomes differ among students in different educational settings, and that these effects canceled each other out in our summarizing analyses. Since students' academic self-concepts in particular were considered as a psychosocial outcome in many studies (e.g., Elbaum, 2002), we examined moderation analyses for self-concept versus other psychosocial factors. We found no differences: Like other psychosocial variables, self-concept did not differ among students with GLD in inclusive versus segregated schools. With regard to self-concept, it is conceivable that the BIRG (students identify themselves with the successes of their class, which increases their own self-concept; e.g., Vogl & Preckel, 2014) and the BFLPE (students develop better self-concepts in lower-performing than in high-performing classes; e.g., Marsh et al., 2019) cancel each other out. Some students may identify themselves with their classmates in terms of performance and develop a higher self-concept and self-esteem in inclusive settings, while others develop lower academic self-concepts as a result of upward comparisons.

Regarding Research Question 6, a funnel plot for studies considering psychosocial outcomes among students with GLD showed an asymmetry, and Egger's test was significant. These findings suggest that a publication bias may exist, as studies with high sampling variance showed higher effect sizes. A further analysis with a  $p$  curve did not indicate  $p$  hacking, as the  $p$  curve was right-skewed, with more high than low significant effect sizes. However, even if there is no evidence for  $p$  hacking, publication bias or other types of bias cannot be ruled out. Future studies should consider possible bias concerning psychosocial outcomes for students with GLD in more detail.

#### *Students Without GLD*

Neither cognitive nor psychosocial outcomes for students without GLD in inclusive schools differed from those of their counterparts in regular schools (Research Questions 3 and 4). A previous meta-analysis showed a slightly positive effect of the inclusion of students with SEN on cognitive outcomes among classmates without any SEN (Szumski et al., 2017). However, this study examined the influence of all types of SEN mingled together and did not investigate psychosocial outcomes. Therefore, comparisons between Szumski et al.'s and our



study are difficult. Due to the dearth of further studies on this issue, interpretations are limited.

For *cognitive outcomes*, it must be taken into account that although the (slightly negative) effect size did not reveal a significant effect of inclusive schooling, only a few studies were considered. For students without GLD, it is possible that more adaptive lessons in inclusive settings due to additional teaching staff (e.g., Dyson et al., 2004) and less attention by teachers (cf. Staub & Peck, 1994) cancel each other out. The study design was also a significant moderator for students without GLD. Increases in students without GLD's performance over time (longitudinally) was the same in inclusive and segregated education, even though students without GLD exhibited lower achievement in inclusive schools than in segregated schools when measured at a single time point. Thus, a kind of selection effect is also possible for students without GLD. Perhaps parents whose students exhibit rather poor performance (but no GLD) are more likely to send their children to inclusive schools so that they can also benefit from the special support and additional teaching staff (e.g., Thuneberg et al., 2013). This might explain the small negative effect of cognitive outcomes measured at a single time point. However, the development of performance over time does not seem to differ between students without GLD in inclusive and segregated schools. A second moderator was the school level: Secondary school students without GLD in inclusive education showed more negative cognitive outcomes than their counterparts in segregated education, while there was no difference between inclusive and segregated education regarding cognitive outcomes in elementary school. This finding could imply that inclusive education has negative effects on cognitive outcomes among students without GLD only after a longer period of time. However, it must be emphasized that only two studies examined elementary education. Thus, the findings can only be interpreted in a very limited way.

Although inclusion does not seem to have overall positive effects on *psychosocial outcomes*, such as self-concept or well-being, among students without GLD, it is nevertheless likely that certain advantages exist. Studies have shown that in accordance with the contact hypothesis, frequent contact to students with SEN can contribute to outcomes such as more positive attitudes toward people with disabilities (e.g., De Boer et al., 2012; Keith et al., 2015; Maras & Brown, 1996). However, these studies could not be included in the present meta-analysis, as they did not refer exclusively to students with GLD or had no control groups.

#### *Implications for Practice*

First of all, the positive effect on cognitive outcomes among students with GLD and the neutral findings on these students' psychosocial outcomes as well as on outcomes among their peers without GLD speak in favor of inclusive schooling. However, implications for further improving inclusive education can also be drawn. First, (prospective) teachers should be prepared for the challenges of an inclusive school system from the very beginning. For example, they should learn how to deal with heterogeneous classes and which teaching methods are useful (e.g., direct instruction, learning strategies, cooperative learning; for an overview, see Mitchell, 2014; Swanson & Hoskyn, 1998). Furthermore, positive attitudes toward students with disabilities are necessary for the successful implementation

of an inclusive school system. Only if teachers are positively inclined toward students with GLD and want to support their students' individual learning requirements good teaching can take place (e.g., Ben-Yehuda et al., 2010; Mazurek & Winzer, 2011; Treder et al., 2000). Second, teachers' competencies should be strengthened to ensure that students in inclusive settings achieve the best possible results. For example, policymakers could provide further funding to recruit additional teaching staff and further training opportunities for teachers (cf. Loreman, 2007; Pijl, 2010). Third, the development and provision of additional support materials for students with GLD, such as digital learning tutorials, should be pursued (cf. Zhang et al., 2015).

#### *Limitations and Future Studies*

The strength of the present study is its high informative value regarding the effects of inclusive educational settings due to its examination of cognitive and psychosocial outcomes among students with GLD as well as their peers without GLD. This meta-analysis suggests that inclusion has generally positive effects on cognitive outcomes for students with GLD and no detrimental effects on psychosocial dimensions for students with and without GLD. Furthermore, to the best of our knowledge, ours is the first study to investigate a possible selection effect in contrast to schooling effects. By using a three-level meta-analytic approach, our study also utilized state-of-the-art methodology.

However, some limitations must be kept in mind. First, we found a high heterogeneity in the primary studies with regard to aspects such as how GLD was diagnosed and how inclusive settings were implemented. We tried to control for this heterogeneity by adding various moderators (e.g., type of diagnosis, study design). Future studies should also include further moderators. For example, the proportion of students with and without GLD in a given classroom or the use of different teaching models might have an impact on students' outcomes in inclusive education (e.g., Szumski et al., 2017).

Second, inclusion is a very heterogeneous field, with participation in inclusive settings ranging from 1 hour per day, for example, to the whole school day. We took a closer look at the extent of inclusion in the primary studies and found that 28 of the 40 studies are based on inclusion of students with GLD for the whole school day. We performed additional analyses with these 28 studies only and found similar results to those involving all 40 studies. However, the exact implementation of inclusion often remains unclear. Furthermore, the number of qualified teaching staff and their level of expertise differ across countries and even across schools within a country. Most primary studies provided no information about the quantitative and personal-related extent of inclusive education. Thus, authors of future studies should carefully explain the conditions of inclusive schooling under study and future meta-analyses should include the level of inclusion as a possible moderator.

Third, GLD is defined differently between and within countries as well as between studies. Therefore, we included all studies focusing on students who experience GLD in several subjects. An IQ ranging between about 50 and 90 was used as an additional criterion (for an overview, see OECD, 2007). However, due to differing criteria and a lack of information about students' diagnoses in the

primary studies, heterogeneity in the students' GLD can be assumed. Nevertheless, by including all studies and their heterogeneities in our meta-analysis, we considered the full range of studies examining outcomes of inclusive schooling for students with GLD. By including a moderator capturing the clarity of diagnosis, we tried to control for the influence of the heterogeneity of GLD on the effects of inclusive education. Since there was no significant difference in effect sizes between the studies with clear and unclear GLD diagnosis, the effect of inclusive education seems to be relatively constant.

One fundamental limitation is the small number of studies that examined outcomes for students without GLD. There is consensus that analyses should be performed only when the number of primary studies is at least  $k = 10$  (e.g., Higgins & Thompson, 2004; Sterne et al., 2011). This requirement could not be met in the analyses of students without GLD. Accordingly, power analyses revealed insufficient power for both cognitive and psychosocial outcomes among students without GLD. Therefore, the findings on cognitive and psychosocial outcomes for students without GLD can be interpreted only in a limited way. Because the number of studies is even smaller when subgroups are formed to analyze potential moderators, the moderation analyses in particular are of limited informative value. For students with GLD, more than 10 studies were available for both cognitive and psychosocial outcomes. However, in the moderation analyses, there were several subgroups with  $k < 10$ . The results of the moderation analyses for students with GLD can therefore also only be interpreted in a limited way for some subgroups (e.g., school type "secondary" for cognitive outcomes or publication status "unpublished" for psychosocial outcomes). Future studies should examine the extent to which inclusion affects students both with GLD and without GLD in order to be able to make more reliable statements. A generally larger number of studies would also increase the  $k$  for subgroups and lead to more interpretable results of moderation analyses.

Furthermore, future studies should specifically investigate selection effects in contrast to schooling effects. For example, studies should examine cognitive variables among students with GLD before the transition to an inclusive versus segregated setting. This would make it possible to investigate a matched sample in terms of previous performance and examine which educational setting results in higher performance gains.

We examined the impact of inclusive education on students. However, in order to assess the full effectiveness of an inclusive school system, teachers' perspectives should also be taken into account. Inclusion makes classes more heterogeneous, resulting in new challenges for teachers (e.g., J. Alston & Kilham, 2004). Future studies should therefore examine what effects an inclusive school system can have on teachers as well.

## Conclusion

The adoption of the CRPD led to the inclusion of an increasing number of students with SEN into regular classrooms. The question of the effects of inclusive schooling compared to segregated schooling is therefore highly relevant. Since a large number of students with SEN have been diagnosed with GLD and different effects of different types of SEN are assumed, we conducted a

meta-analysis of primary studies on inclusion focusing on students with GLD. In addition to students with GLD showing better performance in inclusive schools, it is noticeable that they benefit from higher participation in society as well (e.g., Farrell, 2000). This advantage is reinforced by the fact that no detrimental effect on these students' psychosocial outcomes as a result of inclusive education was found. Furthermore, no detrimental effects on cognitive and psychosocial outcomes among their peers without GLD were found. Therefore, our study leads to a cautious positive conclusion. There is at least no reason why parents should not send their children to inclusive schools.

### ORCID iDs

Sonja Krämer  <https://orcid.org/0000-0002-1509-5917>

Friederike Zimmermann  <https://orcid.org/0000-0002-0552-7050>

### Notes

We thank Kristina Siewert, Peer Ole Gries, Jasmin Lehrmann, and Louisa Kleesiek for assistance in the study selection process and data coding. We thank Keri Hartman for language editing.

The research reported in this article is supported by the project "Lehramt mit Perspektive" (LeaP@CAU; 01JA1623, 01JA1923). This project is part of the program "Qualitätsoffensive Lehrerbildung," a joint initiative of the German Federal Government and the federal states funded by the Federal Ministry of Education and Research. The authors are responsible for the content of this publication.

### References

*References marked with an asterisk indicate studies included in the meta-analysis.*

- Allodi, M. W. (2000). Self-concept in children receiving special support at school. *European Journal of Special Needs Education, 15*(1), 69–78. <https://doi.org/10.1080/088562500361718>
- Allport, G. W. (1954). *The nature of prejudice*. Addison Wesley.
- Alston, J., & Kilham, C. (2004). Adaptive education for students with special needs in the inclusive classroom. *Australasian Journal of Early Childhood, 29*(3), 24–33. <https://doi.org/10.1177/183693910402900305>
- Alston, P., Parker, S., & Seymour, J. (1992). *Children, rights and the law*. Clarendon Press.
- Americans With Disabilities Act of 1990, 42 U.S.C. § 12101 et seq. (1990). <https://www.ada.gov/pubs/adastatute08.htm>
- \*Arampatzi, A., Mouratidou, K., Evaggelinou, C., Koidou, E., & Barkoukis, V. (2011). Social developmental parameters in primary schools: Inclusive settings' and gender differences on pupils' aggressive and social insecure behaviour and their attitudes towards disability. *International Journal of Special Education, 26*(2), 58–69. <https://eric.ed.gov/?id=EJ937175>
- \*Bakker, J. T. A., Denessen, A., Bosmann, A. M. T., Krijger, E. M., & Bouts, L. (2007). Sociometric status and self-image of children with specific and general learning disabilities in Dutch general and special education classes. *Learning Disability Quarterly, 30*(1), 47–62. <https://doi.org/10.2307/30035515>

- Banks, J., & McCoy, S. (2011). *A study on the prevalence of special educational needs*. National Council for Special Education. <https://www.esri.ie/publications/a-study-on-the-prevalence-of-special-educational-needs>
- Bear, G. G., Minke, K. M., & Manning, M. A. (2002). Self-concept of students with learning disabilities: A meta-analysis. *School Psychology Review*, 31(3), 405–427.
- Becherer, J., Köller, O., & Zimmermann, F. (2020). Externalizing behaviour, task-focused behaviour, and academic achievement: An indirect relation? *British Journal of Educational Psychology*. Advance online publication. <https://doi.org/10.1111/bjep.12347>
- Ben-Yehuda, S., Leyser, Y., & Last, U. (2010). Teacher educational beliefs and socio-metric status of special educational needs (SEN) students in inclusive classrooms. *International Journal of Inclusive Education*, 14(1), 17–34. <https://doi.org/10.1080/13603110802327339>
- Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher Education*, 53(4), 483–516. <https://doi.org/10.1007/s10734-005-5628-5>
- \*Bless, G., & Klaghofer, R. (1991). Begabte Schüler in Integrationsklassen: Untersuchung zur Entwicklung von Schulleistungen, sozialen und emotionalen Faktoren. [Gifted students in integration classes: Research on the development of performance, social, and emotional factors]. *Zeitschrift für Pädagogik*, 37(2), 215–223.
- Bless, G., & Mohr, K. (2007). Die Effekte von Sonderschulunterricht und Gemeinsamen Unterricht auf die Entwicklung von Kindern mit Lernbehinderung [The effects of special education and joint teaching on the development of children with learning disabilities]. In J. Walter & F. Wember (Eds.), *Sonderpädagogik des Lernens* (pp. 385–392). Hogrefe.
- \*Brewton, S. (2005). *The effects of inclusion on mathematics achievement of general education students in middle school* [Doctoral dissertation, Seton Hall University]. <https://scholarship.shu.edu/cgi/viewcontent.cgi?referer=https://scholar.google.de/&httpsredir=1&article=2556&context=dissertations>
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279–301. [https://doi.org/10.1207/s15324818ame0704\\_2](https://doi.org/10.1207/s15324818ame0704_2)
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- \*Cardona, C. (1997, March). *Including students with learning disabilities in mainstream classes: A 2-year Spanish study using a collaborative approach to intervention* [Paper presentation]. Annual Meeting of the American Educational Research Association, Chicago, IL. <https://eric.ed.gov/?id=ED407786>
- Carlberg, C., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education*, 14(3), 295–309. <https://doi.org/10.1177/002246698001400304>
- Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley.
- Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, 34(3), 366–375. <https://doi.org/10.1037/0022-3514.34.3.366>

- \*Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental Retardation*, 42(2), 136–144. [https://doi.org/10.1352/0047-6765\(2004\)42%3C136:APOSAI%3E2.0.CO;2](https://doi.org/10.1352/0047-6765(2004)42%3C136:APOSAI%3E2.0.CO;2)
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Congressional Printing Office. <https://eric.ed.gov/?id=ED012275>
- Cooc, N. (2019). Do teachers spend less time teaching in classrooms with students with special educational needs? Trends from international data. *Educational Researcher*, 48(5), 273–286. <https://doi.org/10.3102/0013189X19852306>
- Craddock, N., & Owen, M. J. (2005). The beginning of the end for the Kraepelinian dichotomy. *British Journal of Psychiatry*, 186(5), 364–366. <https://doi.org/10.1192/bjp.186.5.364>
- Daniel, L. G., & King, D. A. (1997). Impact of inclusion education on academic achievement, student behavior and self-esteem, and parental attitudes. *Journal of Educational Research*, 91(2), 67–80. <https://doi.org/10.1080/00220679709597524>
- Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 23(3), 357–374. <https://doi.org/10.3102/00028312023003357>
- De Boer, A., Pijl, S. J., & Minnaert, A. (2012). Students' attitudes towards peers with disabilities: A review of the literature. *International Journal of Disability, Development and Education*, 59(4), 379–392. <https://doi.org/10.1080/1034912X.2012.723944>
- \*Dessemontet, R. S., Bless, G., & Morin, D. (2012). Effects of inclusion on the academic achievement and adaptive behaviour of children with intellectual disabilities. *Journal of Intellectual Disability Research*, 56(6), 579–587. <https://doi.org/10.1111/j.1365-2788.2011.01497.x>
- Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology & Education Quarterly*, 35(1), 75–98. <https://doi.org/10.1525/aeq.2004.35.1.75>
- Dyson, A., Farrell, P., Polat, F., Hutcheson, G., & Gallannaugh, F. (2004). *Inclusion and pupil achievement*. Department for Education and Skills. [https://www.researchgate.net/publication/228647795\\_Inclusion\\_and\\_Pupil\\_Achievement](https://www.researchgate.net/publication/228647795_Inclusion_and_Pupil_Achievement)
- Elbaum, B. (2002). The self-concept of students with learning disabilities: A meta-analysis of comparisons across different placements. *Learning Disabilities: Research & Practice*, 17(4), 216–226. <https://doi.org/10.1111/1540-5826.00047>
- European Agency for Special Needs and Inclusive Education. (2017). *European Agency Statistics on Inclusive Education: 2014 Dataset Cross-Country Report*. <https://www.european-agency.org/resources/publications/european-agency-statistics-inclusive-education-2014-dataset-cross-country>
- European Agency for Special Needs and Inclusive Education. (2018). *European Agency Statistics on Inclusive Education: 2016 Dataset Cross-Country Report*. <https://www.european-agency.org/resources/publications/european-agency-statistics-inclusive-education-2016-dataset-cross-country>
- Farrell, P. (2000). The impact of research on developments in inclusive education. *International Journal of Inclusive Education*, 4(2), 153–162. <https://doi.org/10.1080/136031100284867>
- Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A. M., Nigg, J. T., & Fair, D. A. (2019). The heterogeneity problem: Approaches to identify psychiatric subtypes.

- Trends in Cognitive Sciences*, 23(7), 584–601. <https://doi.org/10.1016/j.tics.2019.03.009>
- Fletcher, J. (2010). Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management*, 29(1), 69–83. <https://doi.org/10.1002/pam.20479>
- Fore, C., Hagan-Burke, S., Burke, M. D., Boon, R. T., & Smith, S. (2008). Academic achievement and class placement in high school: Do students with learning disabilities achieve more in one class placement than another? *Education and Treatment of Children*, 31(1), 55–72. <https://doi.org/10.1353/etc.0.0018>
- Friesen, J., Hickey, R., & Krauth, B. (2010). Disabled peers and academic achievement. *Education Finance and Policy*, 5(3), 317–348. [https://doi.org/10.1162/EDFP\\_a\\_00003](https://doi.org/10.1162/EDFP_a_00003)
- \*Fruth, J. D., & Woods, M. N. (2015). Academic performance of students without disabilities in the inclusive environment. *Education*, 135(3), 351–361.
- Gallegos, J., Langley, A., & Villegas, D. (2012). Anxiety, depression, and coping skills among Mexican school children: A comparison of students with and without learning disabilities. *Learning Disability Quarterly*, 35(1), 54–61. <https://doi.org/10.1177/0731948711428772>
- \*Gorges, J., Neumann, P., Wild, E., Stranghöner, D., & Lütje-Klose, B. (2018). Reciprocal effects between self-concept of ability and performance: A longitudinal study of children with learning disabilities in inclusive versus exclusive elementary education. *Learning and Individual Differences*, 61(January), 11–20. <https://doi.org/10.1016/j.lindif.2017.11.005>
- Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 63–84). Routledge.
- Gresham, F. M., & MacMillan, D. L. (1997). Social competence and affective characteristics of students with mild disabilities. *Review of Educational Research*, 67(4), 377–415. <https://doi.org/10.3102/00346543067004377>
- \*Haeblerlin, U., Bless, G., Moser, U., & Klaghofer, R. (1990). *Die Integration von Lernbehinderten. Versuche, Theorien, Forschungen, Enttäuschungen, Hoffnungen* [The integration of learning disabled: Attempts, theories, research, disappointment, hopes]. Haupt.
- Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, 15(2), 177–199. <https://doi.org/10.1076/sesi.15.2.177.30432>
- \*Hartfield, L. R. (2009). *Mathematics achievement of regular education students by placement in inclusion and non-inclusion classrooms and their principals' perceptions of inclusion* [Doctoral dissertation, University of Southern Mississippi]. <https://aquila.usm.edu/dissertations/1084/>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- \*Hessels, M. G., & Schwab, S. (2015). Mathematics performance and metacognitive behaviors during task resolution and the self-concept of students with and without learning disabilities in secondary education. *Insights on Learning Disabilities: From Prevailing Theories to Validated Practices*, 12(1), 57–72. [https://www.researchgate.net/publication/277313851\\_Mathematics\\_performance\\_and\\_metacognitive\\_behaviors\\_during\\_task\\_resolution\\_and\\_the\\_self-concept\\_of\\_students\\_with\\_and\\_without\\_learning\\_disabilities\\_in\\_secondary\\_education](https://www.researchgate.net/publication/277313851_Mathematics_performance_and_metacognitive_behaviors_during_task_resolution_and_the_self-concept_of_students_with_and_without_learning_disabilities_in_secondary_education)

- \*Hienonen, N., Lintuvuori, M., Jahnukainen, M., Hotulainen, R., & Vainikainen, M. P. (2018). The effect of class composition on cross-curricular competences: Students with special educational needs in regular classes in lower secondary education. *Learning and Instruction, 58*(December), 80–87. <https://doi.org/10.1016/j.learninstruc.2018.05.005>
- Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine, 23*(11), 1663–1682. <https://doi.org/10.1002/sim.1752>
- Hocutt, A. M. (1996). Effectiveness of special education: Is placement the critical factor? *Future of Children, 6*(1), 77–102. <https://doi.org/10.2307/1602495>
- Hornstra, L., Denessen, E., Bakker, J., Van Den Bergh, L., & Voeten, M. (2010). Teacher attitudes toward dyslexia: Effects on teacher expectations and the academic achievement of students with dyslexia. *Journal of Learning Disabilities, 43*(6), 515–529. <https://doi.org/10.1177/0022219409355479>
- Huber, K. D., Rosenfeld, J. G., & Fiorello, C. A. (2001). The differential impact of inclusion and inclusive practices on high, average and low-achieving general education students. *Psychology in the Schools, 38*(6), 497–504. <https://doi.org/10.1002/pits.1038>
- IGI Global. (n.d.). What is cognitive learning outcomes. <https://www.igi-global.com/dictionary/learning-outcomes-across-instructional-delivery/4225>
- Jacobs, N. (2011). Understanding school choice: Location as a determinant of charter school racial, economic, and linguistic segregation. *Education and Urban Society, 45*(5), 459–482. <https://doi.org/10.1177/0013124511413388>
- Justice, L. M., Logan, J. A., Lin, T. J., & Kaderavek, J. N. (2014). Peer effects in early childhood education: Testing the assumptions of special-education inclusion. *Psychological Science, 25*(9), 1722–1729. <https://doi.org/10.1177/0956797614538978>
- Kalambouka, A., Farrell, P., Dyson, A., & Kaplan, I. (2007). The impact of placing pupils with special educational needs in mainstream schools on the achievement of their peers. *Educational Research, 49*(4), 365–382. <https://doi.org/10.1080/00131880701717222>
- Keith, J. M., Bennetto, L., & Rogge, R. D. (2015). The relationship between contact and attitudes: Reducing prejudice toward individuals with intellectual and developmental disabilities. *Research in Developmental Disabilities, 47*(December), 14–26. <https://doi.org/10.1016/j.ridd.2015.07.032>
- Kennedy, C. H., Shukla, S., & Fryxell, D. (1997). Comparing the effects of educational placement on the social relationships of intermediate school students with severe disabilities. *Exceptional Children, 64*(1), 31–47. <https://doi.org/10.1177/001440299706400103>
- Klicpera, C., & Gasteiger-Klicpera, B. (2006). Einfluss des Besuchs einer Integrationsklasse auf die längerfristige Entwicklung von Schülern ohne Behinderung [Influence of attending an integration class on the long-term development of students without disabilities]. *Heilpädagogik, 1*, 11–20.
- \*Kocaj, A., Kuhl, P., Haag, N., Kohrt, P., & Stanat, P. (2017). Schulische Kompetenzen und schulische Motivation von Kindern mit sonderpädagogischem Förderbedarf an Förderschulen und an allgemeinen Schulen [School competencies and school motivation of children with special educational needs in special schools and regular school]. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich, & N. Haag (Eds.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (pp. 302–315). Waxmann.



- \*Kocaj, A., Kuhl, P., Jansen, M., Pant, H. A., & Stanat, P. (2018). Educational placement and achievement motivation of students with special educational needs. *Contemporary Educational Psychology, 55*(October), 63–83. <https://doi.org/10.1016/j.cedpsych.2018.09.004>
- \*Kocaj, A., Kuhl, P., Kroth, A. J., Pant, H. A., & Stanat, P. (2014). Wo lernen Kinder mit sonderpädagogischem Förderbedarf besser? Ein Vergleich schulischer Kompetenzen zwischen Regel- und Förderschulen in der Primarstufe [Where do students with special educational needs learn better? A comparison of achievement between regular primary schools and special schools]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 66*(2), 165–191. <https://doi.org/10.1007/s11577-014-0253-x>
- Kölm, J., Gresch, C., & Haag, N. (2017). Hintergrundmerkmale von Schülerinnen und Schülern mit sonderpädagogischen Förderbedarfen an Förderschulen und an allgemeinen Schulen [Background characteristics of students with special educational needs at special and general schools]. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich, & N. Haag (Eds.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (pp. 291–301). Waxmann.
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(2), 311–328. <https://doi.org/10.1037/0021-9010.78.2.311>
- Kuhl, P., Kocaj, A., & Stanat, P. (2020). Zusammenhänge zwischen einem gemeinsamen Unterricht und kognitiven und non-kognitiven Outcomes von Kindern ohne sonderpädagogischen Förderbedarf [Associations between joint education and cognitive and non-cognitive outcomes of students without special educational needs]. *Zeitschrift für Pädagogische Psychologie*. Advance online publication. <https://doi.org/10.1024/1010-0652/a000283>
- Lackaye, T., Margalit, M., Ziv, O., & Ziman, T. (2006). Comparisons of self-efficacy, mood, effort, and hope between students with learning disabilities and their non-LD-matched peers. *Learning Disabilities: Research & Practice, 21*(2), 111–121. <https://doi.org/10.1111/j.1540-5826.2006.00211.x>
- Learning Disability Association of New York. (n.d.). *Learning disabilities vs. differences*. <http://www.ldanys.org/index.php?s=2&b=25>
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Quarterly, 31*(3), 313–327. <https://doi.org/10.1080/02796015.2002.12086158>
- Lipnevich, A. A., & Roberts, R. D. (2014). Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences, 22*(2), 173–177. <https://doi.org/10.1016/j.lindif.2011.11.016>
- Loreman, T. (2007). Seven pillars of support for inclusive education: Moving from “why?” to “how?” *International Journal of Whole Schooling, 3*(2), 22–38. <https://files.eric.ed.gov/fulltext/EJ847475.pdf>
- Madden, N. A., & Slavin, R. E. (1983). Mainstreaming students with mild handicaps: Academic and social outcomes. *Review of Educational Research, 53*(4), 519–569. <https://doi.org/10.3102/00346543053004519>
- Maras, P., & Brown, R. (1996). Effects of contact on children’s attitudes toward disability: A longitudinal study. *Journal of Applied Social Psychology, 26*(23), 2113–2134. <https://doi.org/10.1111/j.1559-1816.1996.tb01790.x>

- Markussen, E. (2004). Special education: Does it help? A study of special education in Norwegian upper secondary schools. *European Journal of Special Needs Education, 19*(1), 33–48. <https://doi.org/10.1080/0885625032000167133>
- Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist, 24*(3), 231–242. <https://doi.org/10.1027/1016-9040/a000332>
- \*Marston, D. (1996). A comparison of inclusion only, pull-out only, and combined service models for students with mild disabilities. *Journal of Special Education, 30*(2), 121–132. <https://doi.org/10.1177/002246699603000201>
- \*Martlew, M., & Hodson, J. (1991). Children with mild learning difficulties in an integrated and in a special school: Comparisons of behaviour, teasing and teachers' attitudes. *British Journal of Educational Psychology, 61*(3), 355–372. <https://doi.org/10.1111/j.2044-8279.1991.tb00992.x>
- Mazurek, K., & Winzer, M. (2011). Teacher attitudes toward inclusive schooling: Themes from the international literature. *Education and Society, 29*(1), 5–25. <https://doi.org/10.7459/ct/29.1.02>
- McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Barmer, A., Nachazal, T., Barnett, M., & Purcell, S. (2019). *The Condition of Education 2019* (NCES 2019-144). National Center for Education Statistics. <https://nces.ed.gov/pubs2019/2019144.pdf>
- Mitchell, D. (2014). *What really works in special and inclusive education: Using evidence-based teaching strategies*. Routledge.
- Möller, J. (2013). Effekte inklusiver Beschulung aus empirischer Sicht [Effects of inclusive schooling from an empirical perspective]. In J. Baumert, V. Masuhr, J. Möller, C. Pluhar, H. E. Tenorth, & R. Werning (Eds.), *Inklusion. Schulmanagement Handbuch: Band 146* (pp. 15–37). Oldenbourg. <http://www.gbv.de/dms/bs/toc/749859628.pdf>
- Möller, J., Streblow, L., & Pohlmann, B. (2009). Achievement and self-concept of students with learning disabilities. *Social Psychology of Education, 12*(1), 113–122. <https://doi.org/10.1007/s11218-008-9065-z>
- \*Morvitz, E., & Motta, R. W. (1992). Predictors of self-esteem: The roles of parent-child perceptions, achievement, and class placement. *Journal of Learning Disabilities, 25*(1), 72–80. <https://doi.org/10.1177/002221949202500111>
- \*Murawski, W. W. (2006). Student outcomes in co-taught secondary English classes: How can we improve? *Reading & Writing Quarterly, 22*(3), 227–247. <https://doi.org/10.1080/10573560500455703>
- Myklebust, J. O. (2007). Diverging paths in upper secondary education: Competence attainment among students with special educational needs. *International Journal of Inclusive Education, 11*(2), 215–231. <https://doi.org/10.1080/13603110500375432>
- Nakken, H., & Pijl, S. J. (2002). Getting along with classmates in regular schools: A review of the effects of integration on the development of social relationships. *International Journal of Inclusive Education, 6*(1), 47–61. <https://doi.org/10.1080/13603110110051386>
- \*Nusser, L., & Wolter, I. (2016). There's plenty more fish in the sea: Das akademische Selbstkonzept von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf Lernen in integrativen und segregierten Schulsettings [There's plenty more fish in the sea: The academic self-concept of students with learning difficulties

- in integrative and segregated educational settings]. *Empirische Pädagogik*, 30(1), 130–143.
- OECD. (2007). *Students with disabilities, learning difficulties, and disadvantages: Policies, statistics, and indicators*. OECD Publishing. <https://doi.org/10.1787/9789264027619-en>
- Ogelman, H. G., & Seçer, Z. (2012). The effect inclusive education practice during preschool has on the peer relations and social skills of 5-6-year olds with typical development. *International Journal of Special Education*, 27(3), 169–175. <https://files.eric.ed.gov/fulltext/EJ1001069.pdf>
- Oh-Young, C., & Filler, J. (2015). A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities. *Research in Developmental Disabilities*, 47(December), 80–92. <https://doi.org/10.1016/j.ridd.2015.08.014>
- Oxford English Dictionary. (n.d.). Psychosocial. In *Oxford English dictionary*. <https://www.oed.com/view/Entry/153937?redirectedFrom=psychosocial#eid>
- Peetsma, T., Vergeer, M., Roeleveld, J., & Karsten, S. (2001). Inclusion in education: Comparing pupils' development in special and regular education. *Educational Review*, 53(2), 125–135. <https://doi.org/10.1080/00131910125044>
- \*Peleg, O. (2011). Social anxiety among Arab adolescents with and without learning disabilities in various educational frameworks. *British Journal of Guidance & Counselling*, 39(2), 161–177. <https://doi.org/10.1080/03069885.2010.547053>
- Peng, P., Lin, X., Ünäl, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, 146(7), 595–634. <https://doi.org/10.1037/bul0000231>
- \*Pescatore, K. D. (2000). *The effect of educational placement on self-concept* [Master's thesis, Rowan University]. <https://rdw.rowan.edu/cgi/viewcontent.cgi?article=2738&context=etd>
- Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, 90(1), 24–46. <https://doi.org/10.3102/0034654319877153>
- Pijl, S. J. (2010). Preparing teachers for inclusive education: Some reflections from the Netherlands. *Journal of Research in Special Educational Needs*, 10(Suppl. 1), 197–201. <https://doi.org/10.1111/j.1471-3802.2010.01165.x>
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, 86(1), 207–236. <https://doi.org/10.3102/0034654315582067>
- \*Porlier, P., Saint-Laurent, L., & Page, P. (1999, April). *Social contexts of secondary classrooms and their effect on social competence and social adjustment of students with learning disabilities* [Paper presentation]. *Annual Meeting of the American Educational Research Association*. Montreal, CA.
- Porrata, J. L. (1997). Preliminary comparison of scores of special education and regular students on the Eysenck Personality Questionnaire for Children. *Psychological Reports*, 80(1), 191–194. <https://doi.org/10.2466/pr0.1997.80.1.191>
- Pyryt, M., & Bosetti, L. (2007). Parental motivation in school choice: Seeking the competitive edge. *Journal of School Choice*, 1(4), 89–108. <https://doi.org/10.1300/15582150802098795>
- \*Rathmann, K., Vockert, T., Bilz, L., Gebhardt, M., & Hurrelmann, K. (2018). Self-rated health and wellbeing among school-aged children with and without special educational needs: Differences between mainstream and special schools. *Research*

- in *Developmental Disabilities*, 81(October), 134–142. <https://doi.org/10.1016/j.ridd.2018.04.021>
- \*Rea, P. J., McLaughlin, V. L., & Walther-Thomas, C. (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children*, 68(2), 203–222. <https://doi.org/10.1177/001440290206800204>
- \*Rossmann, P., Klicpera, B. G., Gebhardt, M., Roloff, C., & Weindl, A. (2011). Zum Selbstkonzept von SchülerInnen mit Sonderpädagogischem Förderbedarf in Sonderschulen und Integrationsklassen: Ein empirisch fundierter Diskussionsbeitrag [On the self-concept of students with special educational needs in special schools and integrative schools: An empirically based discussion contribution]. In R. Mikula & H. Kittl-Satran (Eds.), *Dimension der Erziehungs- und Bildungswissenschaft* (pp. 107–120). Leykam.
- Ruijs, N. M., & Peetsma, T. T. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, 4(2), 67–79. <https://doi.org/10.1016/j.edurev.2009.02.002>
- Ruijs, N. M., Van der Veen, I., & Peetsma, T. T. (2010). Inclusive education and students without special educational needs. *Educational Research*, 52(4), 351–390. <https://doi.org/10.1080/00131881.2010.524749>
- Salend, S. J., & Duhaney, L. M. (1999). The impact of inclusion on students with and without disabilities and their educators. *Remedial and Special Education*, 20(2), 114–126. <https://doi.org/10.1177/074193259902000209>
- \*Sauer, S., Ide, S., & Borchert, J. (2007). Zum Selbstkonzept von Schülerinnen und Schülern an Förderschulen und in integrativer Beschulung: Eine Vergleichsuntersuchung [On the self-concept of students at special schools and in integrative schooling: A comparative study]. *Heilpädagogische Forschung*, 33, 135–142.
- Schalock, R. L., Luckasson, R. A., & Shogren, K. A. (2007). The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities*, 45(2), 116–124. [https://doi.org/10.1352/1934-9556\(2007\)45\[116:TROMRU\]2.0.CO;2](https://doi.org/10.1352/1934-9556(2007)45[116:TROMRU]2.0.CO;2)
- \*Schmidt, M. (2000). Social integration of students with learning disabilities. *Developmental Disabilities Bulletin*, 28(2), 19–26.
- \*Schwab, S. (2014). Haben sie wirklich ein anderes Selbstkonzept? Ein empirischer Vergleich von Schülern mit und ohne sonderpädagogischem Förderbedarf im Bereich Lernen [Do they really have a different self-concept? An empirical comparison of students with and without learning difficulties]. *Zeitschrift für Heilpädagogik*, 65(3), 116–121.
- \*Schwab, S. (2015). Social dimensions of inclusion in education of 4th and 7th grade pupils in inclusive and regular classes: Outcomes from Austria. *Research in Developmental Disabilities*, 43–44(August–September), 72–79. <https://doi.org/10.1016/j.ridd.2015.06.005>
- \*Schwab, S. (2018). Who intends to learn and who intends to leave? The intention to leave education early among students from inclusive and regular classes in primary and secondary schools. *School Effectiveness and School Improvement*, 29(4), 573–589. <https://doi.org/10.1080/09243453.2018.1481871>
- Schwab, S., Rossmann, P., Tanzer, N., Hagn, J., Oitzinger, S., Thurner, V., & Wimberger, T. (2015). Schulisches Wohlbefinden von SchülerInnen mit und ohne sonderpädagogischen Förderbedarf. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 43(4), 265–274. <https://doi.org/10.1024/1422-4917/a000363>

- \*Sharpe, M. N., York, J. L., & Knight, J. (1994). Effects of inclusion on the academic performance of classmates without disabilities: A preliminary study. *Remedial and Special Education, 15*(5), 281–287. <https://doi.org/10.1177/074193259401500503>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. <https://doi.org/10.1177/1745691614553988>
- Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43–69. <https://doi.org/10.1006/ceps.1996.0004>
- \*Smogorzewska, J., Szumski, G., & Grygiel, P. (2019). Theory of mind development in school environment: A case of children with mild intellectual disability learning in inclusive and special education classrooms. *Journal of Applied Research in Intellectual Disabilities, 32*(5), 1241–1254. <https://doi.org/10.1111/jar.12616>
- Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science, 24*(5), 358–362. <https://doi.org/10.1177/09637214155589345>
- Spencer, T. D. (2013). Self-contained classroom. In F. R. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders* (pp. 109–139). Springer.
- Staub, D., & Peck, C. A. (1994). What are the outcomes for non-disabled students? *Educational Leadership, 52*(4), 36–40.
- \*Stelling, S. (2018). *Schulisches Wohlbefinden von Kindern mit sonderpädagogischem Förderbedarf Lernen. Eine vergleichende Analyse in inklusiven Klassen und Förderschulklassen des dritten und vierten Jahrgangs* [School well-being of children with learning difficulties: A comparative analysis in inclusive and special classes of the third and fourth grade] [Doctoral dissertation, University of Bielefeld]. [https://pub.uni-bielefeld.de/download/2917002/2917003/Dissertation\\_Silke\\_Stelling\\_Schulisches\\_Wohlbefinden.pdf](https://pub.uni-bielefeld.de/download/2917002/2917003/Dissertation_Silke_Stelling_Schulisches_Wohlbefinden.pdf)
- Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal, 343*, Article d4002. <https://doi.org/10.1136/bmj.d4002>
- \*Stranghöner, D., Hollmann, J., Otterpohl, N., Wild, E., Lütje-Klose, B., & Schwinger, M. (2017). Inklusion versus Exklusion: Schulsetting und Lese-Rechtschreibentwicklung von Kindern mit Förderschwerpunkt Lernen [Inclusion vs. exclusion: School Setting and the development of reading and writing skills in students with mild learning difficulties]. *Zeitschrift für Pädagogische Psychologie, 31*(2), 125–136. <https://doi.org/10.1024/1010-0652/a000202>
- Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68*(3), 277–321. <https://doi.org/10.3102/00346543068003277>
- Szumski, G., & Karwowski, M. (2012). School achievement of children with intellectual disability: The role of socioeconomic status, placement, and parents' engagement. *Research in Developmental Disabilities, 33*(5), 1615–1625. <https://doi.org/10.1016/j.ridd.2012.03.030>
- \*Szumski, G., & Karwowski, M. (2014). Psychosocial functioning and school achievement of children with mild intellectual disability in Polish special, integrative, and

- mainstream schools. *Journal of Policy and Practice in Intellectual Disabilities*, 11(2), 99–108. <https://doi.org/10.1111/jppi.12076>
- \*Szumski, G., & Karwowski, M. (2015). Emotional and social integration and the big-fish-little-pond effect among students with and without disabilities. *Learning and Individual Differences*, 43(October), 63–74. <https://doi.org/10.1016/j.lindif.2015.08.037>
- Szumski, G., Smogorzewska, J., & Karwowski, M. (2017). Academic achievement of students without special educational needs in inclusive classrooms: A meta-analysis. *Educational Research Review*, 21(June), 33–54. <https://doi.org/10.1016/j.edurev.2017.02.004>
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37(5), 483–504. [https://doi.org/10.1016/S0883-0355\(03\)00016-8](https://doi.org/10.1016/S0883-0355(03)00016-8)
- Thuneberg, H., Vainikainen, M.-P., Ahtiainen, R., Lintuvuori, M., Salo, K., & Hautamäki, J. (2013). Education is special for all: The Finnish support model. *Gemeinsam Leben*, 21(2), 67–78. [https://www.researchgate.net/publication/268486371\\_Education\\_is\\_special\\_for\\_all\\_-\\_The\\_Finnish\\_support\\_model](https://www.researchgate.net/publication/268486371_Education_is_special_for_all_-_The_Finnish_support_model)
- Törmänen, M. R., & Roebers, C. M. (2018). Developmental outcomes of children in classes for special educational needs: Results from a longitudinal study. *Journal of Research in Special Educational Needs*, 18(2), 83–93. <https://doi.org/10.1111/1471-3802.12395>
- Treder, D. W., Morse, W. C., & Ferron, J. M. (2000). The relationship between teacher effectiveness and teacher attitudes toward issues related to inclusion. *Teacher Education and Special Education*, 23(3), 202–210. <https://doi.org/10.1177/088840640002300303>
- U.K. Public General Acts. (2014). *Children and Families Act, Section 20*. <http://www.legislation.gov.uk/ukpga/2014/6/section/20/enacted>
- United Nations. (1989). *Convention on the Rights of the Child (CRC)*. [https://ec.europa.eu/anti-trafficking/legislation-and-case-law-international-legislation-united-nations/united-nations-convention-rights\\_en](https://ec.europa.eu/anti-trafficking/legislation-and-case-law-international-legislation-united-nations/united-nations-convention-rights_en)
- United Nations. (2006). *Convention on the Rights of Persons with Disabilities*. <http://www.refworld.org/docid/4680cd212.html>
- Vasquez, A. C., Patall, E. A., Fong, C. J., Corrigan, A. S., & Pine, L. (2016). Parent autonomy support, academic achievement, and psychosocial functioning: A meta-analysis of research. *Educational Psychology Review*, 28(3), 605–644. <https://doi.org/10.1007/s10648-015-9329-z>
- Vogl, K., & Preckel, F. (2014). Full-time ability grouping of gifted students: Impacts on social self-concept and school-related attitudes. *Gifted Child Quarterly*, 58(1), 51–68. <https://doi.org/10.1177/0016986213513795>
- \*Waldron, N. L., & McLeskey, J. (1998). The effects of an inclusive school program on students with mild and severe learning disabilities. *Exceptional Children*, 64(3), 395–405. <https://doi.org/10.1177/001440299806400308>
- Wang, M. C., & Baker, E. T. (1985). Mainstreaming programs: Design features and effects. *Journal of Special Education*, 19(4), 503–521. <https://doi.org/10.1177/002246698501900412>
- \*Wild, E., Schwinger, M., Lütje-Klose, B., Yotyodying, S., Gorges, J., Stranghöner, D., Neumann, P., Serke, B., & Kurmitzki, S. (2015). Schülerinnen und Schüler mit dem Förderschwerpunkt Lernen in inklusiven und exklusiven Förderarrangements: Erste Befunde des BiLieF-Projektes zu Leistung, sozialer Integration, Motivation und

- Wohlbefinden [Students with learning disabilities in inclusive and exclusive school settings: First results from the BiLief project on achievement, social integration, motivation, and well-being]. *Unterrichtswissenschaft*, 43(1), 7–21. [https://www.researchgate.net/publication/272530311\\_Schulerinnen\\_und\\_Schuler\\_mit\\_dem\\_Forderschwerpunkt\\_Lernen\\_in\\_inklusiven\\_und\\_exklusiven\\_Forderrangements\\_Erste\\_Befunde\\_des\\_BiLief-Projektes\\_zu\\_Leistung\\_sozialer\\_Integration\\_Motivation\\_und\\_Wohlbefinde](https://www.researchgate.net/publication/272530311_Schulerinnen_und_Schuler_mit_dem_Forderschwerpunkt_Lernen_in_inklusiven_und_exklusiven_Forderrangements_Erste_Befunde_des_BiLief-Projektes_zu_Leistung_sozialer_Integration_Motivation_und_Wohlbefinde)
- World Health Organization. (2004). *ICD-10: International statistical classification of diseases and related health problems* (10th revision, 2nd ed.). World Health Organization.
- Zhang, M., Trussell, R. P., Gallegos, B., & Asam, R. R. (2015). Using math apps for improving student learning: An exploratory study in an inclusive fourth grade classroom. *TechTrends*, 59(2), 32–39. <https://doi.org/10.1007/s11528-015-0837-y>

### Authors

SONJA KRÄMER is a doctoral candidate at Institute for Psychology of Learning and Instruction, Kiel University, Olshausenstraße 75, 24118 Kiel, Germany; email: [skraemer@ipl.uni-kiel.de](mailto:skraemer@ipl.uni-kiel.de). Her research program deals with inclusion/heterogeneity in school and students' social behavior.

JENS MÖLLER is a full professor of educational psychology at Institute for Psychology of Learning and Instruction, Kiel University, Olshausenstraße 75, 24118 Kiel, Germany; email: [jmoeller@ipl.uni-kiel.de](mailto:jmoeller@ipl.uni-kiel.de). His research program deals with academic self-concept, teacher competencies, and bilingual learning.

FRIEDERIKE ZIMMERMANN is a full professor of educational and psychological assessment and inclusion/heterogeneity at Institute for Psychology of Learning and Instruction, Kiel University, Olshausenstraße 75, 24118 Kiel, Germany; email: [fzimmermann@ipl.uni-kiel.de](mailto:fzimmermann@ipl.uni-kiel.de). Her research program deals with inclusion/heterogeneity in school, students' social behavior, teachers' professional development, and academic self-concept.