



Does Professional Development Effectively Support the Implementation of Inclusive Education? A Meta-Analysis

Johanna L. Donath¹ · Timo Lüke^{2,3} · Elisabeth Graf¹ · Ulrich S. Tran⁴ · Thomas Götz¹

Accepted: 21 February 2023 / Published online: 7 March 2023
© The Author(s) 2023

Abstract

Inclusive education is a reform aimed at educating all students in general classrooms, independent of diversity features such as special educational needs, giftedness, or migration. Its successful implementation requires teachers with professional knowledge about inclusive education, skills to address the diverse needs in the classroom, and positive beliefs toward inclusive education. Teachers are provided with professional development opportunities, but are these effective in improving their learning process and positively impacting students' behavior? We conducted a meta-analysis to address this question. The screening of 12,050 search results revealed 342 eligible studies with more than 155,000 participants and 1123 effects from four outcome categories: teachers' knowledge, skills, and beliefs and students' behavior. We observed positive, though varying, effects on all four outcome categories: large effects on teachers' knowledge regarding inclusive education ($g=0.93$), moderate effects on teachers' skills ($g=0.49$), small effects on teachers' beliefs ($g=0.23$), and small-to-moderate effects on student behavior ($g=0.37$). We also examined factors that might explain the differences in the strength of training effects. The data suggest that long-term training with high practical relevance and active learning opportunities facilitates transfer to schools.

Keywords Inclusive education · Professional development · Meta-analysis · Teacher training · Teacher education

✉ Johanna L. Donath
johanna.donath@univie.ac.at

- ¹ Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria
- ² Institute of Education Research and Teacher Education, University of Graz, Graz, Austria
- ³ Research Center for Inclusive Education, Graz, Austria
- ⁴ Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

Introduction

In past decades, education systems have considerably changed how students with diverse educational needs are educated (Alzahrani, 2020). As evidenced by an increasing numbers of countries introducing laws facilitating inclusive education (Allan, 2021), school systems are adapting to address the demands of their diverse communities. Approaches that involve educating students together in general classrooms as opposed to separated classrooms are becoming increasingly popular. Such an approach is thought to overcome diversity-related disadvantages independent of students' various features such as special educational needs, giftedness, or multicultural background (UNESCO, 1994; United Nations, 2005, 2006). Well-prepared and educated teachers are crucial to realize this development toward more inclusive education.

Teachers play a crucial role in implementing inclusive education, as they shape learning opportunities and experiences in school for students (Hattie, 2009) and are responsible for introducing innovations in education systems. To realize educational reforms, teachers should be equipped with professional knowledge, skills for implementation, and positive beliefs about the reform (Bransford et al., 2005; Lipowsky & Rzejak, 2015). The expectation is that students' behavior, achievement, and attitudes will improve if teachers acquire new skills in professional development and successfully apply them in the classroom (Desimone, 2009; Lipowsky & Rzejak, 2015; Pit-Ten Cate et al., 2018).

Introduction and Problem Statement

To better prepare preservice teachers for their future professional life in inclusive classrooms, inclusive education has become obligatory in teacher education programs across many countries. These programs aim to provide knowledge about inclusive education, develop inclusive teaching practices (e.g., collaborative teaching, individualized instruction), and facilitate positive attitudes (Florian & Camedda, 2020). In-service teachers' professional development opportunities often support their ongoing implementation of inclusive education. It should be noted that we distinguish between training pre- and in-service teachers because the circumstances of learning differ (Girardet, 2018; Savolainen et al., 2012), i.e., in-service teachers have the chance to apply newly acquired methods in their classrooms but may also face the challenges with implementing new methods added to their everyday teaching tasks. In turn, preservice teachers can devote more time to learning new teaching methods but have to apply them in hypothetical or exercise contexts. Numerous studies report differences in pre- and in-service teachers' willingness to teach diverse classrooms, attitudes, and concerns (e.g., Edwards et al., 2006; Kimanen et al., 2019; Rumalutur & Kurniawati, 2019), suggesting that the two groups should be looked at separately. We therefore focus specifically on the learning of in-service teachers as they are influencing current practices in schools.

Professional development opportunities are offered to in-service teachers and commonly refer to group training sessions where teachers from different institutions

come together and receive input on a topic. However, the implementation and ongoing support thereof that is supposed to naturally follow such a professional development course are not integrated into these programs (Cramer et al., 2019; Darling-Hammond et al., 2017). Although following this practice equips many teachers with professional development at comparably low costs (Darling-Hammond et al., 2017), its effectiveness in bringing about intended changes in teacher practice can be questioned (Copur-Gencturk & Papakonstantinou, 2016; Garet et al., 2016). Educational research has identified certain aspects that can enhance the efficacy of professional development (Desimone, 2009; Dunst et al., 2015; Sims & Fletcher-Wood, 2021). Content focus, coherence with other learning activities, opportunities for intensive active learning, collective participation of teachers from one institution, and longer duration of training are characteristics that have all been associated with supporting teachers' learning in professional development and modifying their teaching practices (Cordingley et al., 2015; Desimone, 2009). However, these design characteristics are difficult to implement in a one-shot event, which reflects the majority of offered professional development programs (Cramer et al., 2019).

Because providing in-service teachers with opportunities for professional development is a common way to prepare them to implement inclusive education, it is crucial to investigate its effectiveness. Studies have reported null effects for professional development participation (e.g., Adhabi, 2018; Alquraini, 2012; Schmidt, 2019), with some even reporting negative influences on in-service teachers' attitudes toward inclusive education (e.g., Edwards et al., 2006; Jäntschi et al., 2015; Sanches-Ferreira et al., 2018). Even studies that were based on interventions that satisfied many of the above-mentioned design criteria for effective professional development report that only a few participants genuinely understood the concept of interest, felt equipped to implement it (e.g., Carew et al., 2019; Forlin et al., 2014), and not necessarily led to improved student behavior (Garet et al., 2016).

Previous Reviews

The number of studies addressing teacher professional development to improve inclusive education practices has risen with an increasing number of countries introducing inclusive education. Previous reviews have investigated research practices in this field, for example, the applied study designs and assessed outcome variables (Van Mieghem et al., 2018; Waitoller & Artiles, 2016), revealing that studies investigating the influence of professional development addressing inclusive education assess different categories of outcome variables. Specifically, teachers' knowledge, skills, and beliefs, as well as students' behavior, academic achievement, and attitudes toward school, are commonly assessed. Reviews report positive influences of professional development participation but include only a few primary studies in their analyses, with considerable variations in reported effect sizes (Avramidis & Norwich, 2002; Brock & Carter, 2017; Dignath et al., 2022). However, these studies focus on specific aspects of diversity (i.e., special educational needs) or isolated outcome measures (e.g., attitude; Dignath et al., 2022, or implementation fidelity; Brock & Carter, 2017). They mainly comprise studies on English-speaking

countries (e.g., Knight & Wiseman, 2005) or exclusively peer-reviewed studies (e.g., Tristani & Bassett-Gunter, 2020; Van Mieghem et al., 2018). As far as we are aware, no existing meta-analysis differentiates between preservice and in-service teachers, although it is plausible that learning processes differ depending on previous experience and the environment (e.g., Edwards et al., 2006; Kimanen et al., 2019). Thus, we decided to conduct a comprehensive meta-analysis on the effects of professional development in supporting in-service teachers for implementing inclusive education.

Indicators for Improved Implementation of Inclusive Education

Appropriate implementation of inclusive education differs based on the specific context and the students and personnel involved (Lindner & Schwab, 2020; Srivastava et al., 2015). Therefore, professional development providers cannot offer a generic implementation guideline. In order to be able to adjust the implementation of inclusive methods to their classrooms, effective professional development should equip teachers with increased (a) knowledge and (b) skills, (c) a positive change in beliefs, and through changes in instruction lead to (d) improved student behavior (Desimone, 2009).

When investigating the effects of professional development, all four categories of outcome variables are relevant and can be expected to influence one another. For example, knowledge about students' educational needs is necessary to identify their specific learning requirements (Demchenko et al., 2021). Teachers must also be aware of diverse teaching approaches to select an appropriate approach for their students (Lindsay et al., 2014). In addition, applying new teaching approaches can improve students' learning behavior and academic achievements, enabling teachers to perceive positive influences on students (Harris et al., 2014; Lauth-Lebens et al., 2016). Observing struggling learners benefit from changes in instruction can support the development of positive beliefs toward inclusive education, which can facilitate the acquirement of relevant knowledge and its implementation through targeted programs (Ewing et al., 2018).

As the implementation of inclusive education can differ based on the context at hand, researchers can also focus on different aspects and can choose different indicators to investigate successful implementation (Van Mieghem et al., 2018; Waitoller & Artiles, 2016). We conflated the subcategories of outcome variables reflecting all the four categories described in the literature: (a) knowledge can be measured via self-reports and knowledge tests; (b) teachers' skills are measured using assessments of implementation quality of inclusive teaching methods (e.g., collaborative teaching, individualized instruction, Universal Design for Learning, Positive Behavioral Support), variations and frequency of teaching methods, and perceived self-efficacy for inclusive education (one's belief in their ability to positively implement inclusive teaching methods and influence diverse students); (c) attitudes and concerns toward inclusive education and perceptions of inclusive teaching methods are assessed to indicate teachers' beliefs toward inclusive education; and as indicators of (d) student behavior, we conflated two subcategories, namely academic achievement (e.g., in

standardized tests, achievement in research tests, school grades) and other student behavior (including, e.g., learning behavior, school attendance, and attitudes toward school). Data on student behavior were collected via teacher surveys, school data, and students' self-reports.

As shown by previous reviews, these four categories are being studied in the context of implementing inclusive education (Van Mieghem et al., 2018; Waitoller & Artiles, 2016). Primary studies investigating the effectiveness of professional development in this context often assess several indicators alongside (see supplement C), and few studies even assess indicators from the four outcome categories described above (Murthy et al., 2019; Schmidt, 2019; Seibert, 2002). However, there is no meta-analysis that examines relevant outcomes of professional development at both the teacher and student levels.

Study Aims

The current meta-analysis aims to address this research gap by investigating the effects of professional development for in-service teachers on four important outcome categories, specifically to determine whether it (a) enhances knowledge about inclusive education, (b) improves skills of in-service teachers, (c) fosters positive beliefs toward inclusive education, and (d) supports students' behavior. For this study, we conceptualize professional development as structured training for in-service teachers, provided in a group setting and offered to facilitate teachers' preparedness for inclusive education. As the implementation of inclusive education is not limited to the classroom teacher, we include studies on general and special education teachers, teaching assistants, and school administration personnel from kindergarten to high school. Furthermore, we do not limit to specific geographical areas because according to the Salamanca statement from the United Nations (United Nations, 2006), all countries should strive toward implementing and improving inclusive education practices. Contrary to previous reviews, we distinguish between pre- and in-service teachers. It is questionable how meaningful conclusions are when interventions with pre- and in-service teachers are treated as equivalent, because the conditions under which learning takes place differ between pre- and in-service teachers (e.g., Edwards et al., 2006; Kimanen et al., 2019). As in-service teachers mainly shape the current implementation of inclusive education, we focus solely on this group.

The current study aims to identify aspects that enhance the effectiveness of professional development in addressing inclusive education. In order to extend the range of information gathered, we include studies with different designs in this meta-analysis (Dignath et al., 2022; Katsarov et al., 2022). The inclusion of different study designs has been discussed in the literature (Mueller et al., 2018), and this practice has become more common practice, as it can help answer the question of underlying effects and analyze biases introduced by the different designs (Price et al., 2004). Specifically, we aim to include cross-sectional studies, which investigate the influence of previous participation in any professional development program regarding the topic of inclusive education on the constructs of interest (Setia, 2016). In addition, we also include

intervention studies, which investigate the influence of one specific professional development program on the constructs of interest (Fink, 1995). Cross-sectional studies have often been excluded from meta-analyses due to their methodological limitations (e.g., inability to determine causal interpretations and investigate behavior over time), although such studies can provide meaningful insight into teachers' professional development. For example, cross-sectional studies can indicate the effects of naturally occurring professional development participation, including participation in short-term events and multiple programs (O'Connor & Sargeant, 2014), compared to more intensive programs (than commonly provided; Cramer et al., 2019), which are often investigated in intervention studies (Waitoller & Artiles, 2016).

Hypotheses

Overall, we expect positive influences of professional development participation on all four categories of outcome variables. We expect smaller effects on student-level outcomes than on teacher-level outcomes (Desimone, 2009) because these are expected to occur through teachers' behavior changes. Regarding the study design, we expect larger effects in intervention studies than in cross-sectional studies, as the former often assesses the program's effect shortly after its completion, while the latter does not control the time between the program and assessment. Additionally, professional development effects assessed in cross-sectional studies can be expected to be rooted in commonly provided short-term one-shot programs compared to intensive programs investigated in intervention studies. When examining professional development intervention studies, we expect effect sizes to be associated with the number of design criteria described by Garet et al. (2016) and Desimone (2009) met by the training. More specifically, we expect content focus, active learning opportunities, coherence with additional learning activities, longer duration of the training, and collective participation with colleagues to enhance the effectiveness of the professional activities. Some professional development programs offer certification after successful participation. We include this as an additional design aspect because receiving certification can enhance motivation and learning effects (Larsen et al., 2008).

Method

The PRISMA guidelines (Salameh et al., 2020) were used to plan and conduct all phases of the meta-analysis to ensure a transparent process (see preregistration <https://osf.io/jyw5z/>). Supplementary information, including data and code, is available via OSF (<https://osf.io/ehjc3/>).

Literature Search

A systematic search of the literature was conducted in April 2021 using the databases PsycINFO, Web of Science, and ProQuest (ERIC, Education Database, Dissertations & Theses) using the following search terms: inclusion OR inclusive

education OR inclusive classroom AND professional development OR teacher training OR workshop OR teacher education AND teacher OR pedagogical staff OR pedagogical personnel OR teaching assistants OR educators AND school OR K-12 OR kindergarten OR preschool OR vocational college. To reduce the likelihood of publication bias, which refers to the overrepresentation of non-null results in published literature due to selective publishing practices, unpublished studies were screened using the Dissertations & Thesis database of ProQuest and conference abstracts in the search process. In addition, journals and conferences relevant to research on inclusive education were manually searched (see supplement A). After removing duplicates, the literature search revealed 12,050 results, of which 253 were derived solely from journals and 81 from the manual search of conference abstracts.

Inclusion Criteria and Screening

We included all publications that met the following criteria: Studies had to (1) measure the impact of professional development, (2) regarding the topic of inclusive education, (3) among in-service teachers and school personnel (e.g., administrative staff, teaching assistants) on (4) knowledge (assessed with knowledge tests and self-ratings), skills (concerning implementation quality, use of inclusive teaching methods, and self-efficacy for inclusive teaching), beliefs (attitudes, concerns, and perceptions of inclusive education), or student behavior (academic achievement, on-task behavior, school attendance, and attitudes toward school) and (5) had been reported as cross-sectional data, pre- versus post-comparisons of participant data, or comparisons with a control group.

Studies were excluded if preservice teachers were investigated (e.g., Lee et al., 2015; Mulvey et al., 2016) or if professional development consisted solely of coaching, referring to training that took place exclusively in a one-to-one setting (e.g., Gorton et al., 2022). If a specific study was described in both a journal article and a dissertation, the dissertation was included in the meta-analysis because data are commonly reported in more detail in dissertations than in journal articles.

Three coders screened the search results. Two raters independently screened a subsample of 550 abstracts. Divergent decisions concerning the inclusion or exclusion of studies were resolved through discussion and arriving at a consensus. Krippendorff's alpha (2011) indicated high inter-rater reliability ($\alpha = 0.912$). The screening identified 947 abstracts that met the inclusion criteria (see Fig. 1). Initially, we also planned to include qualitative studies that met the mentioned criteria. However, due to the large number of studies identified in the screening process, we decided to exclude the qualitative studies after the screening process. From the 947 studies, 281 qualitative studies were excluded, 65 studies did not fulfill the inclusion criteria when we examined the full text, 209 studies did not report sufficient data for the meta-analysis (e.g., only reported accurate data one time during data collection, did not report accurate data for the control group). Additionally, 27 studies were identified as duplicates and therefore excluded. The authors of studies for which full texts were not available ($k = 34$) were contacted. Thus, we obtained 11 further studies. The remaining 23 studies had to be excluded. Thus, 342 studies were finally included in

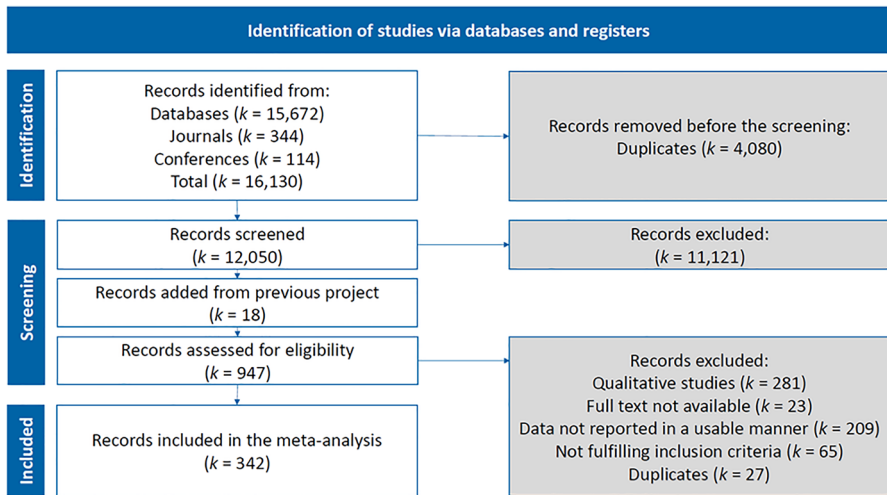


Fig. 1 PRISMA flowchart representing study selection

the meta-analysis. Of these, 14 studies were published in languages other than English (Bulgarian $k = 1$, French $k = 1$, German $k = 3$, Italian $k = 2$, Russian $k = 1$, Spanish $k = 5$, and Turkish $k = 1$). Studies in German and Spanish were directly coded, and studies in other languages were coded with the support of translation software.

Coding of Moderator and Control Variables

To ensure transparency of the coding, a codebook was developed in advance and adjusted during a test phase until all coders were proficient with it (see Supplement B). A pre-configured *M.S. Excel* table was used for the final coding process. The publications identified in the screening process were then coded based on the characteristics of the source of information (e.g., type of publication, country) and study design (e.g., sample size, sampling procedure), participant attributes (e.g., profession, experience with inclusive education), design of the professional development program (e.g., duration, topic, practice opportunities), obtained results (e.g., type of data collection, instrument, analysis method), and type of outcome measure (knowledge, skills, beliefs, student behavior). Again, a subsample of 38 studies was independently coded by two coders. Discrepancies in coding decisions were resolved through discussion and after arriving at a consensus. Inter-rater reliability was high on average (mean Krippendorff's alpha across all coded items $M_\alpha = 0.903$, $SD = 0.088$, $Min = 0.529$, $Max = 1$).

Professional development programs were rated based on design criteria as suggested by Desimone (2009) and Dunst et al. (2015): duration, content focus, coherence with other learning activities, active learning, and collective participation. We used the number of contact hours of the program to determine *duration*. We rated *content focus*, *coherence*, *active learning*, and *collective participation* drawing on the information provided in the studies on scales comprising three criteria, with

scores ranging from 0 to 3. A score of 0 implied that the program did not meet any criterion, a score of 1 implied that the program met one criterion, a score of 2 implied that two criteria were met, and a score of 3 implied that the program met all criteria. The scale *content focus* was rated from 0=general topic to 3=addressing a specific topic (e.g., a specific type of special educational needs) within a specific subject and inclusive teaching method. The *active learning* scale was rated from 0=input session to 3=when the program provided practice opportunities within the sessions, used case studies for explanations and practice, and provided alternating input and praxis phases (compared to blocked design). On the *coherence* scale, a score of 3 was given when implementation was planned in the professional development session, additional coaching was provided, and teachers had to fulfill specific prerequisites to participate in the program (mainly referring to having at least one child with the diversity feature addressed in the program in their classroom). *Collective participation* was rated by the number of colleagues participating in the program, with 0=single participation, 1=participation with at least one colleague, 2=participation with the class team, and 3=participation of (almost) all school staff.

Effect Size Calculation

Cohen's *ds* were chosen as effect sizes and calculated from reported means and standard deviations (Lipsey & Wilson, 2001). If these data were not provided, reported test statistics were used to calculate the effect sizes. The correction factor *J* was applied to correct for the small-sample bias of *d* (Borenstein et al., 2021), resulting in Hedge's *g* (1981) as the effect size metric of the current meta-analysis. The R package *esc* (Lüdtke, 2019) was used to calculate the effect sizes. Effect sizes larger than 2 in absolute value were assumed to be outliers that arose from reporting mistakes. In such cases, we contacted the study authors; if the authors did not respond, effects were replaced by estimates that were two standard deviations from the mean in the respective outcome categories (see Lipsey, 2009; Tukey, 1977). In total, 14 effects were thus replaced.

Summary Effects and Heterogeneity Tests

Summary effects were estimated for each outcome category separately. Effect sizes of $g=0.2$ will be interpreted as small effects, $g=0.5$ as medium effects, and $g=0.8$ as large effects (Cohen, 1977). We expected high heterogeneity in the data, so all analyses were performed with random-effects models. We applied multi-level analyses (Assink & Wibbelink, 2016) to account for the dependency of effect sizes within studies. The (observed) sampling variance of the effect sizes was modeled on level 1, the within-study variance on level 2, and the between-study variance on level 3. The *Q*-test (Borenstein et al., 2021) was used to assess effect-size heterogeneity. Significant values indicated the presence of heterogeneity and suggested conducting moderator analyses to identify potential effect moderators (see below). All analyses were conducted in R with the *metafor* package (Viechtbauer, 2010), and

visualizations were plotted using the *metaviz* package (Kossmeier et al., 2019). Significance was set to $p < 0.05$ (two-tailed).

Moderator Analyses

Moderator analyses were conducted for each outcome category in the case of effect size heterogeneity using the following variables: (1) indicators of study quality (explained in detail below); (2) study characteristics (i.e., intervention vs. cross-sectional study, publication year, years since the legal introduction of inclusive education, continent); (3) data collection (i.e., time between the last session of the professional development program and post-data collection, type of measurement [e.g., observation, questionnaire, vignette] and whether the data collection instrument focused on a specific diversity feature or method); (4) participant characteristics (mean age and teaching experience, school type, percentage of those with experience implementing inclusive education); and (5) professional development design (content focus, active learning, coherence, duration, collective participation, certification). Differences based on the design of programs are analyzed only in intervention studies. We applied the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) to control the false-discovery rate when conducting multiple comparisons for each group of tested moderators addressing the same research question. We used a false-discovery rate of 10% for this procedure.

Assessment of Risk of Bias and Publication Bias

To assess the risk of bias in individual studies, we adapted the Medical Education Research Study Quality Instrument (MERSQI; Reed et al., 2007) for our meta-analysis. The MERSQI has shown high inter-rater reliability and assesses the risk of bias based on quality assessments for each outcome. Our adapted version assessed the (1) study design, (2) response rate, (3) sampling procedure, (4) allocation to conditions, (5) type of data, (6) use of a standardized instrument, (7) internal structure of the instrument, and (8) whether the handling of missing data was reported (see Supplement C). As suggested by the PRISMA guidelines, we analyzed the risk of study bias for each category separately through moderator analyses.

As suggested in the literature, several methods were applied to identify and estimate the risk of publication bias (Rothstein et al., 2005; Siegel et al., 2021). Most methods are difficult to apply with multi-level data. We chose four methods to examine typical sources of publication bias: First, to reduce the risk of publication bias in the first place, unpublished studies were explicitly included in the meta-analyses, and moderator analyses were conducted to test whether published and unpublished studies differed in their effects. Second, contour-enhanced funnel plots (Peters et al., 2008) were created for each outcome category and all their subcategories. In the absence of publication bias, the effect sizes should be symmetrically distributed around the mean effect size, typically in the form of an inverted funnel. Third, the Egger regression test for multi-level data, which regresses the effect sizes on their precision (standard errors) to test for small-study effects (Fernández-Castilla et al.,

2021), was used to support the visual analyses. Effect-size asymmetry in the funnel plot results in a significant Egger regression test if large enough. Fourth, power-enhanced funnel plots were created to include information on the power of individual studies to detect the estimated effect size in the present sample (Kossmeier et al., 2020).

Results

Description of Included Studies

In total, 342 studies met our inclusion criteria (see Supplementary D and O for a complete list). In two cases, two publications were based on the same data but reported different outcome measures and were thus treated as one study (Chao et al., 2016, 2017; Machů, 2015; Machů & Lukeš, 2019).

Studies that met the inclusion criteria were conducted on all continents, with the vast majority were conducted in North America ($k=205$), followed by Europe ($k=68$), Asia ($k=45$), Africa ($k=10$), South America ($k=10$), and Australia ($k=4$). Of the included studies, there were 158 journal articles, 166 dissertations, 12 project reports, and 6 conference papers. In total, 158,713 participants were included in the primary studies, including 62,729 students, 77,787 teachers, and 18,197 members from class and school teams including different professions, such as teaching assistants, and administrative staff. Sample sizes in the primary studies ranged from 3 to 31,000 participants, with a median of 77 participants. Twenty-three studies were conducted with preschool teachers, 115 with primary school teachers, and 76 with secondary school teachers; 128 studies did not define the school type where teachers were employed. Moreover, 188 studies applied a cross-sectional design, 99 had a single-group pre-posttest design, 37 had an independent group pre-posttest design, and 18 had an independent group posttest design. Most studies focused on inclusive education for students with special educational needs ($k=288$), with 76 focusing on specific special educational needs (e.g., autism, learning disabilities). The remaining 54 studies focused on other diversity features, such as, second language learners and gifted students, or addressed multiple categories of heterogeneity.

The professional development programs investigated in the 154 intervention studies ranged from 2 to 750 h, with a median of 20 h, and lasted between half a day and 3 school years, with a median of 3 months. Most programs addressed a specific topic ($k=112$; primarily specific types of special educational needs) but usually did not target a specific subject ($k=126$). Most intervention studies assessed the professional development program's impact immediately after its end ($k=98$); 24 programs offered a certificate to the participants after completing the program, and 51 offered coaching in addition to the training.

In total, 1123 effect sizes were calculated and distributed as follows among the outcome categories: 88 effect sizes for knowledge, 371 for skills, 461 for assessed beliefs regarding inclusive education, and 203 for influences on student behavior (Fig. 2). No differences were observed for effect sizes calculated from means and

standard deviations compared to those calculated from reported test statistics (all $F < 3.8$, all $ps > 0.05$, see Supplement G).

Summary Effects

We calculated summary effects for each outcome category and investigated whether the subcategories within a category differ from each other. We observed significant positive effects in all four outcome categories (Fig. 3). The analysis of knowledge showed a large effect ($g = 0.93$ [0.76; 1.10]), with no difference between self-rated knowledge ($g = 0.96$ [0.70; 1.22]) and knowledge assessed using tests ($g = 0.91$ [0.68; 1.15], $F(1, 86) = 0.50$, $p = 0.48$). A moderate effect was observed on skills to implement inclusive education ($g = 0.49$ [0.41; 0.56]) and on its subcategories (see Fig. 3). The subcategories (implementation quality, use of inclusive methods, self-efficacy for inclusive teaching) did not differ from each other ($F(2, 368) = 0.15$, $p = 0.86$).

We observed a small but positive and significant effect on beliefs toward inclusive education ($g = 0.23$ [0.17; 0.28]), again with no differences between its subcategories ($F(2, 457) = 2.34$, $p = 0.10$). Still, for attitude ($g = 0.23$ [0.18; 0.28]) and perception of inclusive teaching methods ($g = 0.27$ [0.16; 0.39]), a positive effect of professional development participation was observed, while no significant effect was observed for concerns about inclusive education ($g = 0.08$ [-0.14; 0.30]). A small-to-moderate effect of teachers' participation in professional development was observed on student behavior ($g = 0.37$ [0.23; 0.51]), with no differences between

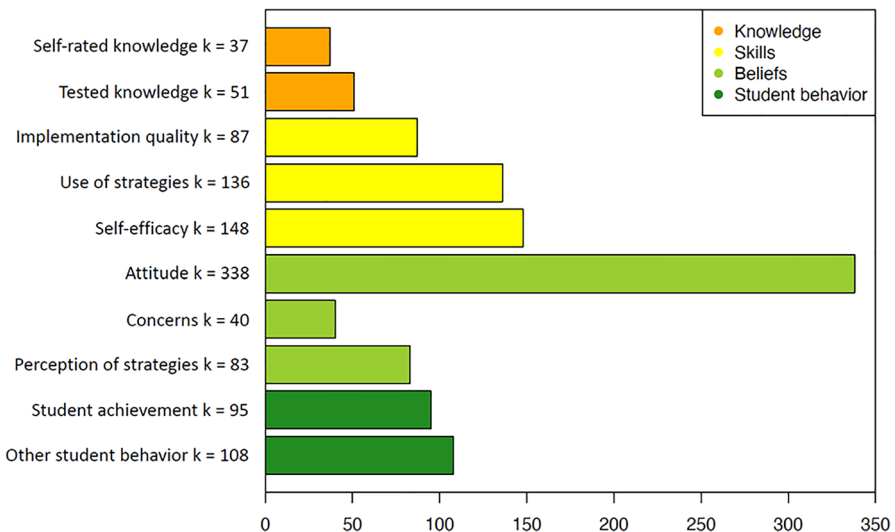


Fig. 2 Distribution of effect sizes by subcategory of outcome categories. Note. k indicates the number of effect sizes in the corresponding outcome category

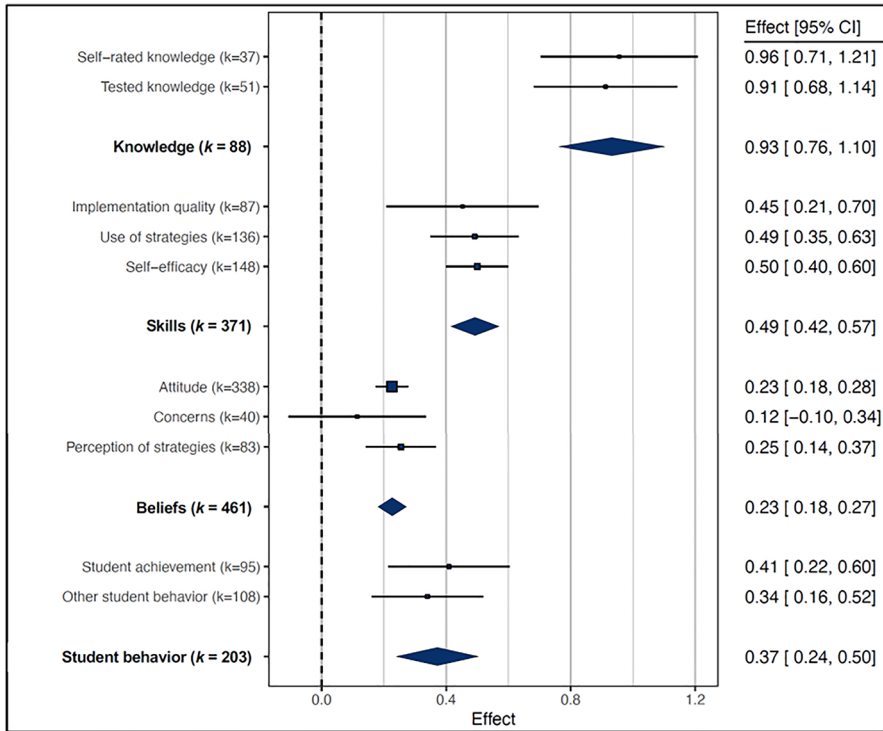


Fig. 3 Overview of summary effects. Note. *k* indicates the number of effect sizes in the corresponding outcome category

student achievement ($g=0.41 [0.22; 0.61]$) and other student behavior ($g=0.29 [0.11; 0.46]$, $F(1, 201)=0.001$, $p=0.98$).

Next, we investigated the presence of heterogeneity for each outcome category, and the *Q*-test was significant for each (Table 1; for more detailed information, see Supplement F). The most variance was found on the between-study level, except for the beliefs category, where variance was mainly present on the within-study level (53.21%). These analyses suggest conducting moderator analyses in all four outcome

Table 1 Overview of summary effects and variance

Outcome category	<i>k</i>	<i>N</i>	Effect size	95% CI	<i>Q</i>	<i>p_Q</i>	Within-study variance (%)	Between-study variance (%)
Knowledge	50	88	0.926	0.76–1.10	1482.96	<.0001	10.24	83.35
Skills	153	371	0.485	0.41–0.56	7446.14	<.0001	34.19	63.76
Beliefs	200	461	0.230	0.18–0.28	2745.3	<.0001	53.21	32.59
Student behavior	51	203	0.372	0.23–0.51	3339.81	<.0001	39.36	58.51

k indicates the number of studies reporting data in the corresponding outcome category and *N* indicates the number of effect sizes per category

categories. The results of moderator analyses across the four outcome categories are summarized in Table 2 (see Supplement K-L for further information), and detailed information on the size and direction of observed effects will be discussed.

Moderator Analyses of Study Characteristics

Publication year, years since the legal anchoring of inclusive education, and the continent where the studies were conducted did not influence the observed effects. Knowledge and beliefs data were not influenced by the control variables describing the study characteristics. Effect sizes reporting effects on skills differed between intervention studies and cross-sectional studies, with the former reporting significantly larger effect sizes ($g=0.56$ [0.46; 0.66]) than the latter ($g=0.36$ [0.27; 0.46]).

Moderator Analyses of Data Collection Characteristics

No differences were observed based on the number of weeks between the last session of the professional development program and post-data collection (Table 2). Effects on knowledge were influenced by the type of instrument used: Studies applying instruments focusing on specific diversity features reported smaller knowledge gains ($g=0.75$ [0.47; 1.02]) compared to instruments that focused on addressing diversity features in inclusive education ($g=1.04$ [0.83; 1.24]). Tested knowledge was larger when assessed with (mainly self-developed) surveys ($g=1.10$ [0.80; 1.39]) than with questionnaires and single-items ($g=0.63$ [0.28; 0.98], $F(2, 48)=3.54$, $p=0.04$).

Effects on skills were not influenced by variables describing the data collection. Regarding beliefs, effects differed based on the applied instruments, with larger effect sizes being observed when instruments focusing on specific teaching methods were used ($g=0.51$ [0.41; 0.60]) compared to instruments focusing on the implementation of inclusive education in general ($g=0.22$ [0.17; 0.27]). The type of measurement influenced data collection on student behavior, with studies applying observational measures reporting larger effects ($g=0.69$ [0.34; 1.03]) than studies using teachers' self-reports ($g=0.16$ [-0.09; 0.40]).

Moderator Analyses of Participant Characteristics

None of the variables describing participant characteristics influenced the observed effect sizes in the categories (Table 2). School type influenced effect sizes assessing the attitudes toward inclusive education subcategory ($F(3, 333)=3.29$, $p=0.02$): We observed positive effects in primary ($g=0.24$ [0.17; 0.32]) and secondary ($g=0.37$ [0.25; 0.48]) school teachers but no effect in kindergarten teachers ($g=-0.03$ [-0.55; 0.49]). Two subcategories—tested knowledge and use of inclusive teaching methods—were influenced by participant characteristics: The higher the mean age, the smaller the effect observed for tested knowledge ($F(1, 23)=8.50$, $p=0.01$, $B=-0.11$, $SE=0.04$), and the more the teachers reported having inclusive teaching

Table 2 Overview of moderator analyses in the outcome categories

Moderator	Knowledge		Skills		Beliefs		Student behavior	
	F	df	F	df	F	df	F	df
Study design								
Publication year	1.06	1, 86	3.22	1, 368	0.08	1, 458	2.56	1, 200
Years since legal anchoring	0.22	1, 86	0.12	1, 368	0.91	1, 458	0.60	1, 200
Intervention studies	3.73	1, 86	8.13	1, 369	0.04	1, 458	0.16	1, 201
Continent	0.67	6, 81	1.03	8, 362	1.60	8, 451	2.27	4, 198
Data collection								
Time between training and data-collection	0.38	1, 75	1.63	1, 257	0.41	1, 210	2.72	1, 194
Instrument Focus: diversity feature	6.57	1, 86	0.00	1, 369	6.21	1, 458	1.08	1, 201
Instrument focus: method	0.03	1, 86	0.86	1, 369	6.34	1, 458	0.15	1, 201
Instrument type	1.99	2, 85	1.40	5, 365	1.57	3, 456	8.53	4, 198
Participant characteristics								
Mean age	1.66	1, 34	0.84	1, 89	0.03	1, 170	0.07	1, 41
Inclusive teaching experience	3.89	1, 39	2.16	1, 214	0.08	1, 281	-	-
School type	0.57	3, 84	1.57	3, 366	2.66	3, 456	2.58	3, 198
Professional development								
Content focus	0.01	1, 75	0.02	1, 257	0.38	1, 210	2.55	1, 194
Active learning	0.00	1, 75	7.84	1, 368	2.14	1, 210	0.02	1, 194
Coherence	0.70	1, 75	2.10	1, 368	0.91	1, 210	1.68	1, 194
Duration	0.02	1, 75	0.56	1, 245	1.05	1, 209	0.01	1, 188
Collective participation	1.89	1, 75	0.07	1, 257	0.85	1, 210	3.86	1, 194
Certification	8.27	1, 75	0.74	1, 257	0.93	1, 209	0.07	1, 194

Degrees of freedom differ based on the information available in the studies

experience, the smaller were effects on the use of inclusive teaching methods ($F(1, 88) = 10.91, p = 0.001, B = -0.02, SE = 0.007$).

Moderator Analyses of Professional Development Design

Content focus did not influence any category of outcome variables, but it did influence a subcategory of student behavior: Changes in student achievement were positively influenced by higher content focus ($F(1, 90) = 5.66, p = 0.02, B = 0.25, SE = 0.11$). *Active learning* was a significant moderator for skills ($F(1, 368) = 7.84, p = 0.005$): Programs with more active learning opportunities reported larger effect sizes ($B = 0.11, SE = 0.04$). Specifically, the indicator alternating versus blocked design explained variance in effects sizes reflecting changes in the use of teaching methods ($F(1, 88) = 9.03, p = 0.004$), with larger changes reported in programs with alternating input and praxis phases ($g = 0.78 [0.57; 0.99]$) than in programs with a blocked design ($g = 0.03 [-0.40; 0.46]$).

Coherence with other learning activities had no influence, except for the subcategory changes in the perception of inclusive teaching methods were positively influenced by additional coherent learning activities ($F(1, 60) = 4.80, p = 0.03, B = 0.11, SE = 0.07$). Analyses of the indicators for coherence showed that programs requiring teachers to fulfill prerequisites for participation reported larger changes in the subcategories of perception of teaching methods ($g = 0.60 [0.33; 0.87]$, $F(1, 59) = 13.58, p = 0.005$) and student achievement ($g = 0.71 [0.18; 1.24]$, $F(1, 90) = 5.20, p = 0.03$) compared to programs that were open to all teachers ($g = 0.13 [0.02; 0.24]$, $g = 0.25 [0.09; 0.42]$, respectively).

We limited the analyses on *duration* to programs lasting up to 200 h, representing about two-thirds of all effect sizes (64.6%), to reduce the influence of extreme programs due to large differences between them (range 2–750 h). Following this reduction, we did not observe influences of the duration of training programs on any of the outcome categories and subcategories (Table 2 and Supplement K-L). *Collective participation* did not influence any outcome category but negatively influenced the subcategory of student achievement ($F(1, 90) = 7.71, p = 0.01, B = -0.21, SE = 0.08$). When all school personnel participated, no effects on student achievement were observed ($g = 0.06 [-0.06; 0.17]$); however, small-to-moderate effects were observed when class teams participated ($g = 0.3 [0.1; 0.5]$) and moderate effects were noted when teachers participated with one colleague ($g = 0.66 [0.16; 1.15]$) and without colleagues ($g = 0.74 [0.1; 1.39]$). Studies offering *certification* after successful completion of the program observed larger effect sizes for knowledge gain ($g = 1.39 [1.07; 1.72]$) than did programs without certification ($g = 0.86 [0.65; 1.08]$), but this did not influence the other categories of outcome variables.

Study Quality

Study bias in the included studies was generally high ($M = 3.74, SD = 1.36, Min = 1, Max = 8.5$). Although moderation analysis of study bias indicated that study quality did not influence the observed effect sizes in the four outcome categories (all

$F < 1.4$, all $p_s > 0.2$, see Supplement H), its influence was observed in the subcategories of self-rated knowledge ($F(1, 35) = 11.06, p = 0.002, B = -0.25, SE = 0.08$) and use of inclusive teaching methods ($F(1, 133) = 4.89, p = 0.03, B = -0.12, SE = 0.05$), where less risk of bias was related to smaller effect sizes.

Publication Bias

Regarding the presence of publication bias, visual analyses of the contour-enhanced funnel plots indicate that the individual effects are roughly symmetrical (see Supplement G, I, and J). Most fall within the 99% confidence interval, and outliers stem from published and unpublished studies. Egger regression tests suggested symmetry for the funnel plots for all outcome categories and subcategories (all $F < 1.3$, all $p_s > 0.2$). The power-enhanced funnel plots (see Fig. 4) illustrate substantial differences in detecting the estimated effect sizes between studies. Moreover, a few studies with very low power were included, but these fall within the normal range of effect sizes. Most studies have low power, especially those assessing beliefs ($med_{power} = 21.5\%$) and student behavior ($med_{power} = 41.8\%$). Power in studies

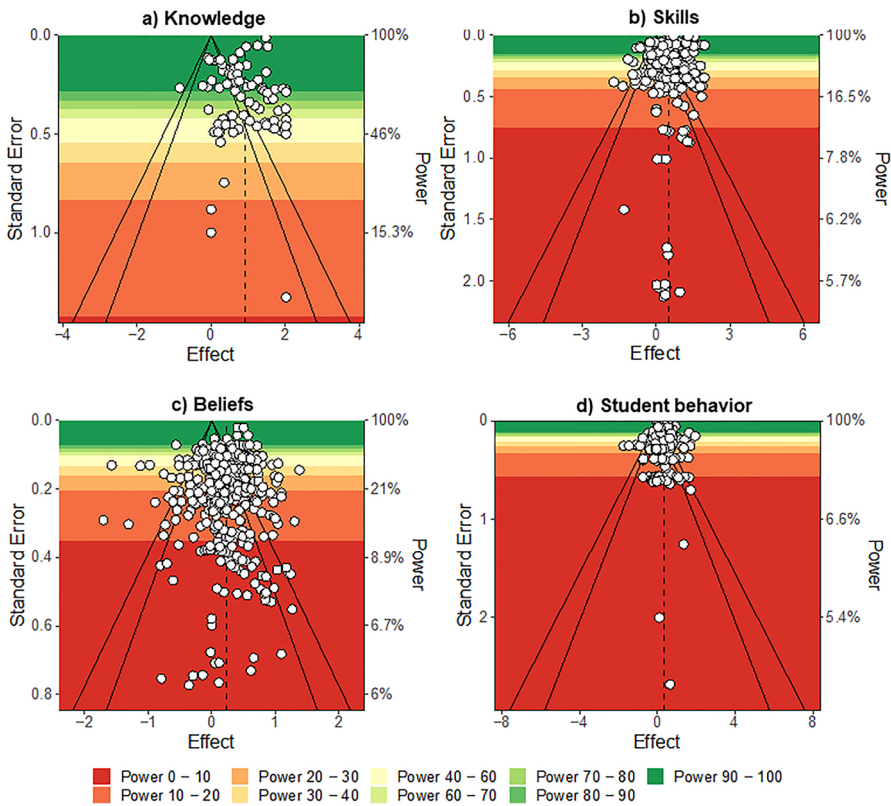


Fig. 4 Power-enhanced funnel plots

assessing skills was rather moderate ($\text{med}_{\text{power}}=69.2\%$) and sufficient in studies assessing knowledge ($\text{med}_{\text{power}}=90.2\%$). Publication status was an inconsequential predictor of knowledge ($F(1, 86)=0.88, p=0.35$) and skills ($F(1, 368)=2.07, p=0.15$), but it moderated effects on beliefs ($F(1, 457)=6.85, p=0.01$) and student behavior ($F(1, 200)=4.77, p=0.03$). Larger effects were reported in published studies (beliefs, $g=0.30$ [0.22; 0.38]; student behavior, $g=0.49$ [0.28; 0.71]) than in unpublished studies ($g=0.19$ [0.13; 0.24], $g=0.17$ [0.07; 0.28], respectively).

These methods suggest that publication bias is present, to a varying degree, in the current meta-analysis in outcome categories and subcategories. The analyses reveal that two typical sources of publication bias do not exert a large influence on the estimated effect sizes, as studies with small sample sizes and low power in this meta-analysis report effect sizes within the normal range. Because half of the included effect sizes stem from unpublished studies (53%), the risk of publication bias was reduced by design.

Discussion

This study aimed to investigate the effectiveness of professional development in supporting the implementation of inclusive education by in-service teachers and analyze design aspects of professional development programs in this regard. A comprehensive meta-analysis was conducted to investigate the effects of professional development on four categories of outcome variables. This meta-analysis is the first to consider different indicators for the effectiveness of professional development addressing inclusive education on both the teacher and student levels together because professional development aims to disseminate knowledge and skills to in-service teachers and help them develop positive attitudes toward the topic in order to improve students' behavior and school experiences (Desimone, 2009).

Through a systematic literature search in five databases, relevant journals, and conferences, we identified 342 studies from more than 50 countries across all continents that assessed the effects of professional development addressing inclusive education on at least one of the outcome variables on in-service teachers and reported quantitative data. As inclusive education is a complex field with varying indicators for good implementation depending on the context and focus, we included outcome variables that reflected the described outcome categories. In sum, we collected 1123 effect sizes spread via the four outcome categories of teachers' knowledge ($k=88$), skills ($k=371$), beliefs ($k=461$), and students' behavior ($k=203$). Overall, we observed the hypothesized positive influence of professional development participation on all four outcome variables. The expected larger effect sizes on the teacher-level outcomes compared to student-level outcomes were observed for teachers' knowledge and skills but not for teachers' beliefs. Contrary to our hypothesis, we observed larger effect sizes in intervention studies compared to cross-sectional studies only in teachers' skills, while no difference was observed in the other outcome categories. Regarding the design principles of professional development, we observed little support for the hypothesized larger effect sizes in programs fulfilling

more criteria of good design compared to programs fulfilling fewer criteria. These results will be discussed in detail in the following sections.

Does Professional Development Improve Teachers' Knowledge, Skills, Beliefs, and Students' Behavior?

The primary approach to improving teaching practices through professional development is disseminating information to in-service teachers. Our study reveals that professional development improves teachers' knowledge, as expressed by the large effect sizes ($g = 0.93$ [0.76; 1.1]). Knowledge gains were smaller among older teachers. This could be due to older teachers perhaps having more existing knowledge and therefore profit less from mere information dissemination or who may be less willing to engage with new information and practices (Saborit et al., 2016; Thomas, 2012).

In sum, assessments using knowledge tests reveal that teachers have more information about inclusive education, while assessments using self-reports show that teachers also perceive themselves as more knowledgeable when participating in professional development. Hence, concerns about the lack of knowledge, which teachers often consider an obstacle to implementing inclusive education, can be addressed via professional development programs.

Another goal of professional development is to improve teachers' skills, and our analysis reveals that professional development considerably improved teachers' skills ($g = 0.49$ [0.41; 0.56]). This finding is in line with Dignath et al. (2022), who observed a large effect ($d = 0.63$) but based their analysis on five studies that mainly included preservice teachers whose confidence in their teaching skills was still developing. Further, the smaller effect observed in the current study may be rooted in methodological issues. In contrast to Dignath and colleagues, the current meta-analysis included studies with control groups to adjust estimated effect sizes for typical developments. We also included unpublished studies and observed that published studies reported larger effect sizes for self-efficacy than unpublished ones, as expected, due to the tendency that positive, statistically significant findings and large effects are more often prioritized by researchers in peer-reviewed studies (i.e., file-drawer-effect, Rosenthal, 1979). Estimated effects in the current study can therefore be interpreted as more realistic. Our study reveals that professional development supports the self-perceived capability to implement inclusive education and actual execution, as implementation quality, variance in the use of teaching methods, and frequency in using evidence-based practices increased. However, improvement does not necessarily mean that the implementation was sufficient, and many practices still need to improve (Nilholm, 2021; Ramberg & Watkins, 2020).

Regarding teacher belief changes, we observed a small positive effect size ($g = 0.23$ [0.17; 0.28]). In detail, we observed this effect on attitudes toward inclusive education and perceptions of teaching methods, while concerns about inclusive education were not influenced by professional development. Professional development aims to provide teachers with information but does not impact the implementation in schools, whereas the hurdles against implementation being faced by teachers in

the schools themselves are often the main focus of teachers' concerns (Abakah et al., 2022; Sharma et al., 2018). Therefore, it is not surprising that concerns about inclusive education are not influenced by professional development. Teachers perceive many obstacles regarding implementing inclusive education; these concerns are likely to inhibit implementation intentions (Miesera et al., 2019) and should therefore be addressed in future programs. In general, changing beliefs requires much effort as it is expected to result from acquiring new knowledge and having positive experiences (Gregoire, 2003). However, the small effects observed for beliefs may also be due to the self-selection of participants in professional development and research, as these participants are expected to have somewhat positive beliefs about the topic of interest in the first place. When teachers had to fulfill specific criteria to participate in the program—usually having a child with a specific type of special educational needs in their classroom—changes regarding the perception of teaching methods were larger compared to programs open to all teachers. This supports the assumption that applying newly acquired knowledge and skills is relevant to change beliefs.

Teachers undergo professional development to improve students' behavior, academic performance, learning behavior, and school experiences. In the current study, we observed a small-to-moderate effect on students' behavior ($g = 0.37$ [0.23; 0.51]). Finding positive effects on the student level is in line with the meta-analysis by Brock and Carter (2017), who observed large effect sizes ($g = 1.08$). However, the programs investigated in the meta-analysis by Brock and Carter (2017) were more intensive than those in the current analysis. They included interventions with pre-service teachers, who are expected to benefit more from professional development programs as they cannot rely on teaching practice (e.g., Kimanen et al., 2019; Lautenbach & Heyder, 2019; Rumalutur & Kurniawati, 2019). In this study, we included different study designs and observed positive effects on students in all designs. Contrary to expectations, effects assessed with objective measures were larger than effects assessed via teacher reports or students' self-reports. This may be caused by teachers' problems in accurately and sensitively identifying changes in students' behavior while teaching that class.

How Should Professional Development Be Designed to Enhance Effectiveness?

Desimone (2009) identified five design criteria that affect the efficacy of professional development: content focus, coherence with other learning activities, opportunities for active learning, collective participation of teachers from one institution, and longer duration. Our study revealed little empirical support for these design criteria, echoing recent finding (Sims & Fletcher-Wood, 2021). Active learning was the only design principle with a significant influence on one outcome category, change in teachers' skills, while the other design principles merely influenced different sub-categories. We analyzed certification after completing the program as an additional design aspect. Not surprisingly, it was found to influence teachers' knowledge gains as people tend to concentrate more on learning provided information when receiving certification (Larsen et al., 2008).

However, the lack of support for these design principles should not be overstated, as descriptions of the programs were scarce. For example, when the description did not provide information on the content, we coded the program as addressing inclusive education in general, although this might not have been the case. Therefore, we expect that design of the programs has more influence on their effectiveness than observed in our study. Further, we lack data from short-term programs, most commonly offered to in-service teachers but rarely addressed by research. Intervention studies primarily focus on assessing the effects of intensive programs, and our moderator analyses are therefore limited to identifying aspects that enhance efficacy within such programs. This might explain the lack of support for influences of program duration. However, we observed some design features that may improve the programs' efficacy:

Although content focus did not influence data on the overall student level, we observed that when the topic addressed in the programs was more specific, students' achievement improved. Coherence with other learning activities improved teachers' perception of inclusive teaching methods. When teachers had to fulfill specific criteria to participate in the program, larger effect sizes were observed in students' achievement. These prerequisites mainly involved teachers having at least one child in their classroom with specific special educational needs addressed in the program. This indicates that teachers apply learned content more easily when they can relate the provided information to their classroom and are probably more motivated because these teachers come with specific questions and needs.

Providing more opportunities for active learning positively affected change in teachers' skills particularly in their use of inclusive teaching methods. The opportunities to apply newly learned content and methods seem to support the development of teaching skills. This effect was influenced by designing the program as a blocked session or alternating input and praxis phases. Skills in general and their use of inclusive teaching methods, in particular, improved when teachers attempted new methods in their classroom, reflected on their process, and received feedback in the next session. This supports the finding of Brock and Carter (2017), who observed that implementation quality improved, especially when teachers had a chance to observe modeling and receive performance feedback.

In our study, collective participation hurt students' achievement. The more teachers from one school participated, the smaller the positive effect on students' achievement. We attribute this to the fact that more teachers are required to participate in the program unlike single participation, which is mainly based on teachers' willingness to do so. This indicates that forcing teachers to participate in professional development can undermine positive changes. It has also been shown that attitudes can be contagious among colleagues (Pedaste et al., 2021). Hence, negative attitudes toward professional development can spread within the collegiate and undermine the implementation intentions of individual teachers.

In summary, intensive professional development programs positively influence teachers' knowledge, skills, and beliefs, as well as students' behavior. The present analysis also provides directions for the design of future professional development programs: Since we assume that effects at the student level occur through improved teaching practices, we suggest providing opportunities for active learning, especially

designing it with alternating input and practical phases. According to our data, asking participating teachers to think of specific students they seek to support can also enhance learning effects.

Limitations

About two-thirds of the integrated works considered in this review were intervention studies. With a median duration of 3 months, the professional development programs investigated in those studies were relatively intensive to programs usually offered to teachers (Cramer et al., 2019). As we expect the intensity of programs to enhance effectiveness, the effect sizes estimated in this study may have overestimated the true effects of professional development. Further, intervention studies tend to measure the program's effects shortly after its end, reflected in a median of zero weeks between the end of the program and post-evaluation. We observed significant positive effects of professional development in both intervention and cross-sectional studies. Differences were observed only for the category of teachers' skills, although we calculated moderate effect sizes for both. Hence, the estimated effects exist but may be overestimated.

Previous reviews investigating the effect of professional development with regard to inclusive education limited their inclusion criteria to one study design. For example, randomized controlled trials, as investigated by Brock and Carter (2017), are more often applied in intensive professional development programs, while more expensive evaluations are applied for expensive programs. Dignath et al. (2022) only included single-group studies whose effects can be overestimated as there is no chance to control for natural development. This assumption is supported by our analyses, as intervention studies tended to report larger effect sizes than did cross-sectional studies, especially single-group studies. Therefore, we chose to include different study designs to not only investigate differences in reported effect sizes but also balance strengths and weaknesses. As mentioned, our analyses indicated significant positive effects for all study designs.

Overestimation of effects can also be rooted in publication bias. As more than half of the included literature was unpublished, the risk of publication bias for our meta-analysis was reduced by design. We did not discover notable indications for the presence of publication bias. The difference between published and unpublished studies was substantiated, yet again confirming the file-drawer effect (Rosenthal, 1979). We observed low power among many included studies. Against assumptions based on publication bias, the effects reported in studies with very low power lay within the range of the remaining studies and reported small effect sizes. Therefore, studies with low power did not inflate the estimated effects. Still, the small number of studies with large power was disappointing.

Implications for Future Research

Concerning the significance of supporting teachers in implementing a more inclusive school system, the results reveal that professional development is a helpful

building block. However, teachers need more support and adapted school frameworks to achieve satisfactory implementation. Moreover, to assess the effectiveness of short-term events commonly offered as professional development, more data from these kinds of programs and their long-term effects are needed. Researchers should not only focus on intensive programs but also investigate real-life learning opportunities for teachers.

Based on our work, we highly recommend that researchers consider statistical power when planning a study. When assessing beliefs or other self-reported variables, we recommend using instruments that generate imaginations of concrete situations, such as including a child with ADHD in the classroom, to reduce the influence of socially desired response behavior. To improve transparency and replicability, providing a more detailed description of the professional development programs and making learning materials easily available are essential, as it would allow more detailed analyses of the program design.

Our study observed positive effects on all outcome categories and subcategories but on concerns about inclusive education. These could not be addressed adequately in such programs. However, since worries are a barrier to implementation and are communicated by teachers, researchers and trainers should examine how to reduce concerns in professional development.

Conclusion

Our study investigated the effectiveness of professional development to improve the implementation of inclusive education, revealing that professional development is a promising strategy to improve not only teachers' knowledge, skills, and beliefs but also students' behavior. This review is the first to investigate the effects of professional development on the teacher and student levels simultaneously. Compared to previous studies, we applied a comprehensive literature search considering different characteristics of professional development programs and research practices, which allowed us to identify a vast number of effect sizes in all four outcome categories and estimate small confidence intervals, reiterating the positive influence of professional development addressing inclusive education. Our analyses show that, in particular, knowledge transfer is effective via professional development. The study findings align with previous reviews and provide new insights regarding the design aspects of professional development programs. The analyses reveal that in-service teachers can and should be supported via professional development to improve their implementation of inclusive education.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-023-09752-2>.

Funding Open access funding provided by University of Vienna.

Data Availability Data and code are available via OSF (<https://osf.io/ehjc/3>).

Declarations

Conflict of Interest The authors declare no competing interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abakah, E., Widin, J., & Ameyaw, E. K. (2022). Continuing professional development (CPD) practices among basic school teachers in the central region of Ghana. *Sage Open*, *12*(2), 1–14. <https://doi.org/10.1177/21582440221094597>
- Adhabi, E. (2018). *The perceptions of elementary school special education and general education teachers on full inclusion of students with autism spectrum disorder (ASD) in Saudi Arabia*. (Publication No. 10814676) [Doctoral dissertation, Saint Louis University]. ProQuest Dissertations and Theses Global.
- Allan, J. (2021). Inclusive education, democracy and COVID-19. A time to rethink? *Utbildning & Demokrati – Tidskrift För Didaktik Och Utbildningspolitik*, *30*(1), 9–21. <https://doi.org/10.48059/uod.v30i1.1549>
- Alquraini, T. A. (2012). Factors related to teachers' attitudes towards the inclusive education of students with severe intellectual disabilities in Riyadh, Saudi. *Journal of Research in Special Educational Needs*, *12*(3), 170–182. <https://doi.org/10.1111/j.1471-3802.2012.01248.x>
- Alzaharani, N. (2020). The development of inclusive education practice: A review of literature. *International Journal of Early Childhood Special Education*, *12*(1), 68–83. <https://doi.org/10.20489/intjces.722380>
- Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. <https://doi.org/10.20982/tqmp.12.3.p154>
- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration / inclusion: A review of the literature. *European Journal of Special Needs Education*, *17*(2), 129–147. <https://doi.org/10.1080/08856250210129056>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.
- Borenstein, M., Hedges, L. V., Higgins, Julian, P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Bransford, J., Darling-Hammond, L., & LePage, P. (2005). *Preparing teachers for a changing world: What teachers should learn and should be able to do*. Jossey-Bass.
- Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education*, *38*(3), 131–144. <https://doi.org/10.1177/0741932516653477>
- Carew, M. T., Deluca, M., Groce, N., & Kett, M. (2019). The impact of an inclusive education intervention on teacher preparedness to educate children with disabilities within the Lakes Region of Kenya. *International Journal of Inclusive Education*, *23*(3), 229–244. <https://doi.org/10.1080/13603116.2018.1430181>
- Chao, C. N. G., Forlin, C., & Ho, F. C. (2016). Improving teachers self-efficacy for teachers in inclusive classrooms in Hong Kong. *International Journal of Inclusive Education*, *20*(11), 1142–1154. <https://doi.org/10.1080/13603116.2016.1155663>

- Chao, C. N. G., Chow, W. S. E., Forlin, C., & Ho, F. C. (2017). Improving teachers' self-efficacy in applying teaching and learning strategies and classroom management to students with special education needs in Hong Kong. *Teaching and Teacher Education*, 66, 360–369. <https://doi.org/10.1016/j.tate.2017.05.004>
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (Revised). Academic Press.
- Copur-Gencturk, Y., & Papakonstantinou, A. (2016). Sustainable changes in teacher practices: A longitudinal analysis of the classroom practices of high school mathematics teachers. *Journal of Mathematics Teacher Education*, 19(6), 575–594. <https://doi.org/10.1007/s10857-015-9310-2>
- Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., & Coe, R. (2015). *Developing great teaching: Lessons from the international reviews into effective professional development*. Teacher Development Trust. <https://dro.dur.ac.uk/15834/1/15834.pdf>
- Cramer, C., Johannmeyer, K., & Drahmann, M. (2019). *Fortbildungen von Lehrerinnen und Lehrern in Baden-Württemberg* [Professional development for teachers in Baden-Württemberg]. Gewerkschaft Erziehung und Wissenschaft. https://www.pedocs.de/volltexte/2019/16567/pdf/Cramer_Johannmeyer_Drahmann_2019_Fortbildungen_von_Lehrerinnen.pdf
- Darling-Hammond, L., Hyster, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/17357/46%20Effective_Teacher_Professional_Development_REPORT.pdf?sequence=1
- Demchenko, I., Maksymchuk, B., Bilan, V., Maksymchuk, I., & Kalynovska, I. (2021). Training future physical education teachers for professional activities under the conditions of inclusive education. *Broad Research in Artificial Intelligence and Neuroscience*, 12(3), 191–213. <https://doi.org/10.18662/brain/12.3/227>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Dignath, C., Rimm-Kaufman, S., van Ewijk, R., & Kunter, M. (2022). Teachers' beliefs about inclusive education and insights on what contributes to those beliefs: A meta-analytical study. *Educational Psychology Review*, 1–53. <https://doi.org/10.1007/s10648-022-09695-0>
- Dunst, C. J., Bruder, M. B., & Hamby, D. W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews*, 10(12), 1731–1744. <https://doi.org/10.5897/err2015.2306>
- Edwards, C. J., Carr, S., & Siegel, W. (2006). Influences of experiences and training on effective teaching practices to meet the needs of diverse learners in schools. *Education*, 126(3), 580–592.
- Ewing, D. L., Monsen, J. J., & Kielblock, S. (2018). Teachers' attitudes towards inclusive education: A critical review of published questionnaires. *Educational Psychology in Practice*, 34(2), 150–165. <https://doi.org/10.1080/02667363.2017.1417822>
- Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *Journal of Experimental Education*, 89(1), 125–144. <https://doi.org/10.1080/00220973.2019.1582470>
- Fink, A. (1995). *Evaluation for education & psychology*. Sage Publications, Inc.
- Florian, L., & Camedda, D. (2020). Enhancing teacher education for inclusion. *European Journal of Teacher Education*, 43(1), 4–8. <https://doi.org/10.1080/02619768.1707579>
- Forlin, C., Loreman, T., & Sharma, U. (2014). A system-wide professional learning approach about inclusion for teachers in Hong Kong. *Asia-Pacific Journal of Teacher Education*, 42(3), 247–260. <https://doi.org/10.1080/1359866X.2014.906564>
- Garet, M. S., Heppen, J., Walters, K., Parkinson, J., Smith, T., Song, M., Barrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development (NCEE 2016–4010)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://files.eric.ed.gov/fulltext/ED569154.pdf>
- Girardet, C. (2018). Why do some teachers change and others don't? A review of studies about factors influencing in-service and pre-service teachers' change in classroom management. *Review of Education*, 6(1), 3–36. <https://doi.org/10.1002/rev3.3104>
- Gorton, K., Allday, R. A., Lane, J. D., & Ault, M. J. (2022). Effects of brief training plus electronic feedback on increasing quantity and intonation of behavior specific praise among pre-school teachers. *Journal of Behavioral Education*, 31, 731–750. <https://doi.org/10.1007/s10864-020-09427-w>

- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, 15(2), 147–179. <https://doi.org/10.1023/A:1023477131081>
- Harris, K. J., Graham, S., & Adkins, M. (2014). Practice-based professional development and self-regulated strategy development for tier 2, at-risk writers in second grade. *Contemporary Educational Psychology*, 40, 5–16. <https://doi.org/10.1016/j.cedpsych.2014.02.003>
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- Jäntschi, C., Jäeuth, J., Bosse, S., & Spörer, N. (2015). Fortbildung und Beratung zum inklusiven Unterricht [Training and consulting on inclusive education]. In *Inklusives Lernen und Lehren im Land Brandenburg Abschlussbericht zur Begleitforschung des Pilotprojekts „Inklusive Grundschule“ [Inclusive Learning and Teaching in the State of Brandenburg Final Report on the Accompanying Research of the Pilot Project "Inclusive Elementary School"]* (pp. 155–180). Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM). https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/schule/schulentwicklung/Evaluation/Evaluationsberichte/2015_PING_Abschlussbericht.pdf#page=157
- Katsarov, J., Andorno, R., Krom, A., & van den Hoven, M. (2022). Effective strategies for research integrity training. *Educational Psychology Review*, 34, 935–955. <https://doi.org/10.1007/s10648-021-09630-9>
- Kimanen, A., Alisaari, J., & Kallioniemi, A. (2019). In-service and pre-service teachers' orientations to linguistic, cultural and worldview diversity. *Journal of Teacher Education and Educators*, 8(1), 35–54.
- Knight, S. L., & Wiseman, D. L. (2005). Professional development for teachers of diverse students: A summary of the research. *Journal of Education for Students Placed at Risk*, 10(4), 387–405. https://doi.org/10.1207/s15327671espr1004_3
- Kossmeier, M., Tran, U. S., & Voracek, M. (2019). Visual inference for the funnel plot in meta-analysis. *Zeitschrift Für Psychologie*, 227(1), 83–89. <https://doi.org/10.1027/2151-2604/a000358>
- Kossmeier, M., Tran, U. S., & Voracek, M. (2020). Power-enhanced funnel plots for meta-analysis: The sunset funnel plot. *Zeitschrift Für Psychologie*, 228(1), 43–49. <https://doi.org/10.1027/2151-2604/a000392>
- Krippendorff, K. (2011). *Computing Krippendorff's Alpha Reliability*. Retrieved from https://repository.upenn.edu/asc_papers/43
- Larsen, D. P., Butler, A. C., & Roediger, H. L., III. (2008). Test-enhanced learning in medical education. *Medical Education*, 42, 959–966. <https://doi.org/10.1111/j.1365-2923.2008.03124.x>
- Lautenbach, F., & Heyder, A. (2019). Changing attitudes to inclusion in preservice teacher education: A systematic review. *Educational Research*, 61(2), 231–253. <https://doi.org/10.1080/00131881.2019.1596035>
- Lauth-Lebens, M., Lauth, G., & Rohrsdorfer, I. (2016). Effects of a training for schoolteachers on behavioural problems of ADHD students and teacher strain. *Verhaltenstherapie & Verhaltensmedizin*, 37(2), 140–153.
- Lee, F. L. M., Yeung, A. S., Tracey, D., & Barker, K. (2015). Inclusion of children with special needs in early childhood education. *Topics in Early Childhood Special Education*, 35(2), 79–88. <https://doi.org/10.1177/0271121414566014>
- Lindner, K. T., & Schwab, S. (2020). Differentiation and individualisation in inclusive education: A systematic review and narrative synthesis. *International Journal of Inclusive Education*, 1–21. <https://doi.org/10.1080/13603116.2020.1813450>
- Lindsay, S., Proulx, M., Scott, H., & Thomson, N. (2014). Exploring teachers' strategies for including children with autism spectrum disorder in mainstream classrooms. *International Journal of Inclusive Education*, 18(2), 101–122. <https://doi.org/10.1080/13603116.2012.758320>
- Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione*, 7(2), 27–51.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, 4, 124–147. <https://doi.org/10.1080/15564880802612573>
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Sage Publications.

- Lüdecke, D. (2019). *R Package 'esc' - Effect Size Computation for Meta Analysis - v0.5.1*. 36. <http://104.40.9.224/web/packages/esc/esc.pdf>
- Machů, E. (2015). Analyzing differentiated instructions in inclusive education of gifted preschoolers. *Social and Behavioral Sciences*, 171(2010), 1147–1155. <https://doi.org/10.1016/j.sbspro.2015.01.224>
- Machů, E., & Lukeš, P. (2019). Teachers' work with taxonomy of educational objectives as one of the forms of the gifted preschoolers' development. *Acta Educationis Generalis*, 9(3), 1–15. <https://doi.org/10.2478/atd-2019-0011>
- Miesera, S., DeVries, J. M., Jungjohann, J., & Gebhardt, M. (2019). Correlation between attitudes, concerns, self-efficacy and teaching intentions in inclusive education evidence from German pre-service teachers using international scales. *Journal of Research in Special Educational Needs*, 19(2), 103–114. <https://doi.org/10.1111/1471-3802.12432>
- Mueller, M., D'Addario, M., Egger, M., Cevallos, M., Dekkers, O., Mugglin, C., & Scott, P. (2018). Methods to systematically review and meta-analyse observational studies: A systematic scoping review of recommendations. *BMC Medical Research Methodology*, 18(1), 1–18. <https://doi.org/10.1186/s12874-018-0495-9>
- Mulvey, B. K., Chiu, J. L., Ghosh, R., & Bell, R. L. (2016). Special education teachers' nature of science instructional experiences. *Journal of Research in Science Teaching*, 53(4), 554–578. <https://doi.org/10.1002/tea.21311>
- Murthy, M. K. S., Rajaram, P., Mudiyanuru, K. S., Marimuthu, P., Govindappa, L., & Dasgupta, M. (2019). Potential for increased epilepsy awareness: Impact of health education program in school on teachers and children. *Journal of Neurosciences in Rural Practice*, 10(4), 625–630. <https://doi.org/10.1055/s-0039-3399473>
- Nilholm, C. (2021). Research about inclusive education in 2020 – How can we improve our theories in order to change practice? *European Journal of Special Needs Education*, 36(3), 358–370. <https://doi.org/10.1080/08856257.2020.1754547>
- O'Connor, A. M., & Sargeant, J. M. (2014). Meta-analyses including data from observational studies. *Preventive Veterinary Medicine*, 113(3), 313–322. <https://doi.org/10.1016/j.prevetmed.2013.10.017>
- Pedaste, M., Leijen, Ä., Kivirand, T., Nelis, P., & Malva, L. (2021). School leaders' vision is the strongest predictor of their attitudes towards inclusive education practice. *International Journal of Inclusive Education*, 1–17. <https://doi.org/10.1080/13603116.2021.1994661>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61, 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Pit-Ten Cate, I. M., Markova, M., Krischler, M., & Krolak-Schwerdt, S. (2018). Promoting inclusive education: The role of teachers' competence and attitudes. *Insights into Learning Disabilities*, 15(1), 49–63.
- Price, D., Jefferson, T., & Demicheli, V. (2004). Methodological issues arising from systematic reviews of the evidence of safety of vaccines. *Vaccine*, 22(15), 2080–2084. <https://doi.org/10.1016/j.vaccine.2004.01.009>
- Ramberg, J., & Watkins, A. (2020). Exploring inclusive education across Europe: Some insights from the European agency statistics on inclusive education. *Forum for International Research in Education*, 6(1), 85–101. <https://doi.org/10.32865/fire202061172>
- Reed, D. A., Cook, D. A., Beckmann, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *Journal of American Medical Association*, 298(9), 1002–1009.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. John Wiley & Sons.
- Rumalutur, N. A., & Kurniawati, F. (2019). The attitudes of in-service and pre-service teachers toward inclusive education. *Advances in Social Science, Education and Humanities Reserach*, 229, 619–630. <https://doi.org/10.2991/iciap-18.2019.53>
- Saborit, J. A. P., Fernández-Río, J., Cecchini Estrada, J. A., Méndez-Giménez, A., & Alonso, D. M. (2016). Teachers' attitude and perception towards cooperative learning implementation: Influence of continuing training. *Teaching and Teacher Education*, 59, 438–445. <https://doi.org/10.1016/j.tate.2016.07.020>

- Salameh, J.-P., Bossuyt, P. M., McGrath, T. A., Thombs, B. D., Hyde, C. J., Macaskill, P., Deeks, J. J., Leeftang, M., Korevaar, D. A., Whiting, P., Takwoingi, Y., Reitsma, J. B., Cohen, J. F., Frank, R. A., Hunt, H. A., Hooft, L., Rutjes, A. W. S., Willis, B. H., Gatsonis, C., ... McInnes, M. D. F. (2020). Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): Explanation, elaboration, and checklist. *Research Methods and Reporting*, 370, 1–17. <https://doi.org/10.1136/bmj.m2632>
- Sanches-Ferreira, M., Lopes-dos-Santos, P., Alves, S., & Silveira-Maia, M. (2018). The use of the ICF-CY for describing dynamic functioning profiles: Outcomes of a teacher training programme applied in Portugal. *International Journal of Inclusive Education*, 22(7), 734–750. <https://doi.org/10.1080/13603116.2017.1412507>
- Savolainen, H., Engelbrecht, P., Nel, M., & Malinen, O. P. (2012). Understanding teachers' attitudes and self-efficacy in inclusive education: Implications for pre-service and in-service teacher education. *European Journal of Special Needs Education*, 27(1), 51–68. <https://doi.org/10.1080/08856257.2011.613603>
- Schmidt, E. M. (2019). *Supporting preschool teachers to promote physically active play among children with autism spectrum disorders*. [Doctoral dissertation, Northeastern University]. Northeastern University Boston. <https://repository.library.northeastern.edu/files/neu:m044kn747/fulltext.pdf>
- Seibert, J. K. (2002). *Enhancing teacher competence in accomodating challenging behaviors: Investigating the effects of an experimental professional development program* (Publication No. 060494) [University of Wisconsin-Madison]. ProQuest Dissertations and Theses Global.
- Setia, M. S. (2016). Methodology series module 3: Cross-sectional studies. *Indian Journal of Dermatology*, 61(3), 261–264.
- Sharma, U., Armstrong, A. C., Merumeru, L., Simi, J., & Yared, H. (2018). Addressing barriers to implementing inclusive education in the Pacific. *International Journal of Inclusive Education*. <https://doi.org/10.1080/13603116.2018.1514751>
- Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2021). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology*, 1–27. <https://doi.org/10.1037/apl0000991>
- Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: A critical review. *School Effectiveness and School Improvement*, 32(1), 47–63. <https://doi.org/10.1080/09243453.2020.1772841>
- Srivastava, M., de Boer, A., & Pijl, S. J. (2015). Inclusive education in developing countries: A closer look at its implementation in the last 10 years. *Educational Review*, 67(2), 179–195. <https://doi.org/10.1080/00131911.2013.847061>
- Thomas, M. S. C. (2012). Brain plasticity and education. *Educational Neuroscience*, 11(8), 143–156.
- Tristani, L., & Bassett-Gunter, R. (2020). Making the grade: Teacher training for inclusive education: A systematic review. *Journal of Research in Special Educational Needs*, 20(3), 246–264. <https://doi.org/10.1111/1471-3802.12483>
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.
- UNESCO. (1994). *Salamanca statement and framework for action on special needs education*. UNESCO.
- United Nations. (2005). *Guidelines for inclusion: Ensuring access to education for all*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000140224>
- United Nations. (2006). *United Nations Convention on the Rights of Persons with Disabilities*. UNESCO. <https://www.bcscs.org/wp-content/uploads/2020/01/CRPD-Position-StatementCPA.pdf>
- Van Mieghem, A., Verschueren, K., Petry, K., & Struyf, E. (2018). An analysis of research on inclusive education: A systematic search and meta review. *International Journal of Inclusive Education*, 24(6), 675–689. <https://doi.org/10.1080/13603116.2018.1482012>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Waitoller, F. R., & Artiles, A. J. (2016). Teacher learning as curating: Becoming inclusive educators in school/university partnerships. *Teaching and Teacher Education*, 59, 360–371. <https://doi.org/10.1016/j.tate.2016.07.007>