The Transfer of Teacher Training to Inclusive Classroom Practice: A Meta-Analytic SEM Approach

Johanna Lowis Donath[1,2] · **Timo Lüke**[3] · **Ulrich S. Tran**[4] · **Elouise Botes**[5] · **Thomas Goetz**[1]

Abstract

Teachers require support to implement inclusive education effectively, and teacher training is a key to providing this support cost-effectively. However, while education research has established theories on student learning, there is still a lack of understanding regarding how teachers learn and integrate new ideas, making it essential to investigate the transfer process to improve the implementation of school reforms. The many overlapping theories on teachers' learning in in-service training, as presented in the literature, have been summarized here in five models. These five models were tested regarding their fit to empirical data in secondary analyses using meta-analytic SEM. The literature search revealed 228 studies that met the inclusion criteria, comprising 244 samples and 117,609 participating teachers and students. Only one model had an acceptable model fit. This model is characterized by the assumption that teacher training affects willingness to change, which in turn influences the subsequent transfer process. Further research should be theory-based, testing theories using primary data, particularly in the context of willingness to change.

✉ Johanna Lowis Donath
johanna.donath@uni-tuebingen.de

[1]   Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

[2]   Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany

[3]   Institute of Education, University of Kassel, Kassel, Germany

[4]   Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

[5]   Department of Behavioral and Cognitive Sciences, Faculty of Humanities, Education and Social Sciences, Luxembourg, Centre for Educational Testing, University of Luxembourg, Esch-Sur-Alzette, Luxembourg

🖄 Springer

Introduction

Schools are constantly confronted with political reforms, technological advancements, and other societal changes that necessitate adjustments in their organization and daily practices. However, it is not a given that principals and teachers will necessarily implement these reforms and changes, and that the systemic changes will result in improvements for students. Therefore, investing in processes of implementing change in school environments is essential to close the 'implementation gap' (Gouëdard et al., 2020). Teachers play a central role in realizing school reforms as they engage with the content of these programs and the school and classroom environments where these programs are expected to be implemented. In such contexts, teachers' agency is situated within complex, multidimensional, and multilevel work environments (Imants & Van der Wal, 2020). The principal tool to support teachers in this highly demanding change process is to offer training (Popova et al., 2022; Scott et al., 2024), however, understanding the effectiveness of training and the mechanisms through which learning occurs and effects improvements in teaching practice and student outcomes is opaque, with many competing theories found in the literature. To effectively research teacher training, it is essential to develop a conceptual understanding of how learning and transfer processes occur in this highly complex environment (Kennedy, 2016). This includes understanding transfer as a continuum that encompasses the processes from the delivery of knowledge and skills during teacher training to their long-term and successful implementation, contributing to improved student outcomes (Grossman & Salas, 2011). This study aims to collect and summarize existing theories of learning transfer in this context and test these theories using empirical data.

Inclusive Education

Inclusive education (IE) refers to a school reform and a teaching approach that impacts all aspects of school life, affecting the entire school environment as well as individual classrooms. Educating all students together and considering the needs of all learners involves all educational staff and goes far beyond simply adding a new method or tool; it fundamentally alters the approach to teaching and learning (Suprayogi et al., 2017). This reform is the subject of global discourse (Somma, 2020) and is particularly challenging as it affects the entire school framework and requires profound, systemic changes. It is not a matter of teachers deciding whether to use a particular tool. Instead, the way teachers implement IE influences the learning of all students. Therefore, well-trained teachers are crucial, not only for students with special needs but for all students, since implementing IE ultimately impacts the entire school climate and the learning process.

Teacher Training for School Reforms

Effective schools equip students with cultural skills and social competencies, enabling all students to live self-determined and independent lives. To get closer

to this goal and adapt school education to social changes, school reforms are decided upon at the political level to initiate widespread change. They are to be implemented by teachers at the local schools. For successful implementation, teachers must be well informed about these reforms, have acquired the necessary skills, and have positive beliefs regarding this school reform (Lipowsky & Rzejak, 2015). This can be achieved by addressing the training of prospective teachers and, more relevant in terms of numbers, by offering further training for in-service teachers to familiarize them with the reforms (Popova et al., 2022). Unfortunately, providing teachers with information about the current reform is insufficient to adequately address the implementation gap (Cohen & Mehta, 2017; Copur-Gencturk & Papakonstantinou, 2016). This also applies to the context of IE, with many training programs reporting little to no improvement in teachers and students (e.g., Loyalka et al., 2019; Scott et al., 2024). However, some studies have found that training teachers for IE can be successful in bringing about change (Chalmers & Gardiner, 2015; Donath et al., 2023).

To create effective training programs, it is essential to identify design characteristics that enhance their effectiveness and understand how learning transfer occurs from training to students' achievement improvements (Clarke & Hollingsworth, 2002). Although training designs can differ widely (Cramer et al., 2019) and many training programs do not fulfill known criteria for effectiveness (Popova et al., 2022), it has been shown that training participation, in general, can bring about development in the desired direction, in teachers as well as in their students (Didion et al., 2020; Donath et al., 2023). However, teacher training is typically investigated only at the teacher level (i.e., assessing teachers' knowledge and skill acquisition; Dreer et al., 2017) and assumes that sustainable changes will occur later (Dunst et al., 2015). This may be due to the difficulties in evaluating reform implementation in individual classrooms, as successful implementation can be highly individualized and may unfold over time (Sasson & Miedijensky, 2023). Nevertheless, one should refrain from assuming that successful implementation can be achieved through mere participation in training courses (Darling-Hammond et al., 2009; Doherty, 2011). These transfer processes require time for teachers to deepen their understanding of reform practices (Guskey & Yoon, 2009). The design and effectiveness of teacher training have been observed to vary significantly more than in education programs in other contexts (Evans & Popova, 2016; McEwan, 2015); yet, the general recognition of the importance of teacher training is not questioned (Kennedy, 2016). In a longitudinal study, Doherty (2011) found that most participating teachers did not apply what they had learned in the training. This necessitates a closer examination of the transfer processes from training participation to improvements in students' school experiences (Castillo et al., 2018; Florian & Camedda, 2020).

Furthermore, teacher trainings are typically held for groups of teachers to reduce costs, facilitate exchange among teachers, and disseminate coherent information to many individual teachers (Darling-Hammond et al., 2017). Although there are similarities between students' learning in schools and professional teachers' learning and change processes, as one person with expertise typically teaches groups of learners, it is inappropriate to apply learning theories from school contexts to professional teachers' learning and change processes (Kennedy, 2016). Due to their expertise in

the professional field, particularly their responsibility for the learning process and individualized implementation (Moon, 2004), learners are subjected to fundamentally different demands. Therefore, it is important to carefully examine the descriptions of learning and transfer processes in models that address professional training.

Models of Learning in Teacher Training

Numerous theories describe the transfer processes from (teacher) training to reform implementation (Desimone, 2023), with no widely accepted framework available (Nīmante et al., 2025). Some attempts have been made to describe overarching frameworks for studying teacher training (e.g., Darling-Hammond et al., 2017; Desimone, 2009; Kennedy, 2005; Sims et al., 2023; Timperley et al., 2008), but these are often primarily concerned with the design and circumstances of the training. Lists of effective design features can be helpful when designing a training program, following the logic that providing a good product will automatically lead to good outcomes (Opfer & Pedder, 2011). However, focusing on structural features, such as a specific number of hours, can distract from identifying relevant steps in learning processes and how to support them best (Sztajn et al., 2011).

Among theoretical considerations on professional learning in in-service training, there is consensus that external and internal factors influence the learning process. These include personal predispositions, such as intelligence and learning motivation, as well as contextual conditions, including the design of teacher training and the provision of external support during the learning process (Nīmante et al., 2025). This support encompasses spatial and material conditions within the school, the role of school leadership, collaboration among staff, and other school resources. However, theories differ in whether and how they incorporate these factors into their models (Desimone, 2023). In the following section, these influencing factors will be omitted to allow a focus on the steps in the transfer process.

Existing theories differ in the number of steps involved in the transfer process and the directional influence of these steps (Boylan et al., 2018; Kennedy, 2016; Sims et al., 2023), as there is a tendency to increasingly delineate the indicators in transfer theories and to analyze increasingly finer stages of the learning process. Furthermore, researchers have often argued that efforts for school reform are continuous processes (e.g., Clarke & Hollingsworth, 2002; Imants & Van der Wal, 2020; Shulman & Shulman, 2004), assuming a cyclical pattern where the individual learning steps are mutually influencing one another (Imants & Van der Wal, 2020). This assumption is supported by the study of Shaha et al. (2015), who found that repeated participation in teacher training has increasingly positive effects on students' learning behavior.

Although numerous studies have employed professional learning models to explore how teacher change occurs within their contexts, often selecting models that require comparably little effort in the evaluation process focusing on change in teachers (Desimone, 2023; Kowalski et al., 2020), only a few have specifically focused on investigating the transfer process itself as a central research topic (Strang,

2022). In this context, transfer refers to the process by which input provided during teacher training is meaningfully applied in classroom practice, ultimately benefiting student learning and development (Grossman & Salas, 2011). Furthermore, many studies evaluating the effects of teacher training do not use transfer models as the basis of their evaluation (Ahadi et al., 2024; Basma & Savage, 2023). There is a broad consensus that a comprehensive theory describing professional learning processes and their application remains lacking (Basma & Savage, 2023; Boylan et al., 2023; Nīmante et al., 2025). The following section will describe five typical models of the learning and transfer process (see Fig. 1 and Table 1).

Knowledge-to-Practice Model

The simplest transfer models assume that knowledge acquisition changes classroom practices, thus improving students' learning outcomes (e.g., Baldwin & Ford, 1988; Koehler & Mishra, 2009; Lieberman & Miller, 1990; Meiers & Ingvarson, 2005; see Fig. 1a and Table 1). This model assumes a serial succession from knowledge acquisition to improved classroom practices and student outcomes, implying successful application of the acquired knowledge (Baldwin & Ford, 1988; Ignatavicius, & Chung, 2016). This model is advantageous in its simplicity. However, in the context of IE, this might be insufficient, as its implementation requires more than applying a new teaching method; it also involves redefining the responsibilities and concepts of successful teaching styles, as well as selecting and adapting the appropriate teaching methods to the classroom conditions. Such underlying changes can be assumed to provoke feelings of being overwhelmed and uncertainty about how to initiate the change

process.

Nevertheless, being informed about IE and inclusive teaching methods is an important step for implementation. Still, studies report that even though teachers have heard about positive effects and are more knowledgeable after the training, many never end up putting what they have learned into practice (Doherty, 2011; Hargreaves, 2005), thus having little impact on students' achievement (Randel et al., 2016). Since knowledge acquisition does not appear to be sufficient to explain its application, transfer models should include additional links between knowledge and teaching practice (Osman & Warner, 2020).

Knowledge-Skills-Beliefs Model

The following more complex models assume that training not only impacts teachers' professional knowledge acquisition but also affects the skills and beliefs of teachers regarding the reform (e.g., Blume et al., 2010; Desimone, 2009; Egert et al., 2018; see Fig. 1b and Table 1). These changes in teachers' knowledge, skills, and beliefs together trigger changes in teaching practices, resulting in positive effects on students (Clarke & Hollingsworth, 2002; Desimone, 2009; Egert et al., 2018). These models are most commonly applied in evaluation research (Sancar et al., 2021), as they often fit intuitively into the current training program being evaluated (e.g.,

Table 1 Overview of the five models of learning transfer from teacher training to the classroom

Model	Central aspect	Strengths	Limitations	Example
Knowledge-to-Practice	Acquisition of knowledge	Simplicity	Acquired knowledge is often not applied and implemented	Marc learned about classroom management in inclusive settings in the teacher training and applied two methods successfully afterwards
Knowledge-Skills-Beliefs	Acquisition of knowledge, skills, and beliefs	Well-established concept Recognizes that implementation requires more than information	Belief change is slow and effortful, unlikely that training alone changes beliefs	Through the training, Susan gained a deeper understanding of the concept and the legal framework surrounding inclusive education, developed a more positive attitude towards its principles, and felt more confident in her ability to implement inclusive teaching methods in her classroom. Equipped with new knowledge, skills, and attitudes, she introduced strategies to improve classroom management, which in turn led to noticeable improvements in student behavior
Reaction-to-Training	Satisfaction with the training	Focus on training likeability	Learning often driven by challenges, not comfort	Janine really liked the training program, she felt comfortable and perceived the trainer as well-informed. That helped her concentrate and motivated her to apply some of the mentioned inclusive teaching methods that focused on classroom management in her classroom, where she observed positive effects on her students' behavior

Table 1 (continued)

Model	Central aspect	Strengths	Limitations	Example
Practice-to-Belief	Positive experience with implementation	Focus on skill acquisition and sustained implementation Recognizes that belief change takes time and effort Overlooks internal factors beyond self-efficacy	Hard to assess long-term, individual change	The training on classroom management, which introduced inclusive strategies and allowed him to practice some methods, increased his confidence in applying them in his own classroom. In the following weeks, the improved learning behavior of his students led to a shift in his attitude towards inclusive education, strengthening his belief that it is indeed possible to teach a diverse group of students in one classroom

Table 1 (continued)

| Model	Central aspect	Strengths	Limitations	Example
Willingness-to-Change	Willingness to change	Recognizes willingness to change as aspect that can be fostered in training, not as prerequisite	Unclear descriptions how acquisition of knowledge, skills beliefs, and change willingness interact Complexity of the model Recent and not yet well-established	In the training on inclusive education, Elaine was confronted with provocative questions that prompted her to reflect on her current teaching practices in light of the newly introduced concepts for implementing classroom management. This led to a shift in her perspective: She became willing to try the new approach in her classroom. Eager to understand how the method could be adapted, she actively engaged with the training content and explored ways to tailor it to her own teaching context. This supported her to adapt one method and apply it in her classroom, that helped improve her students' behavior
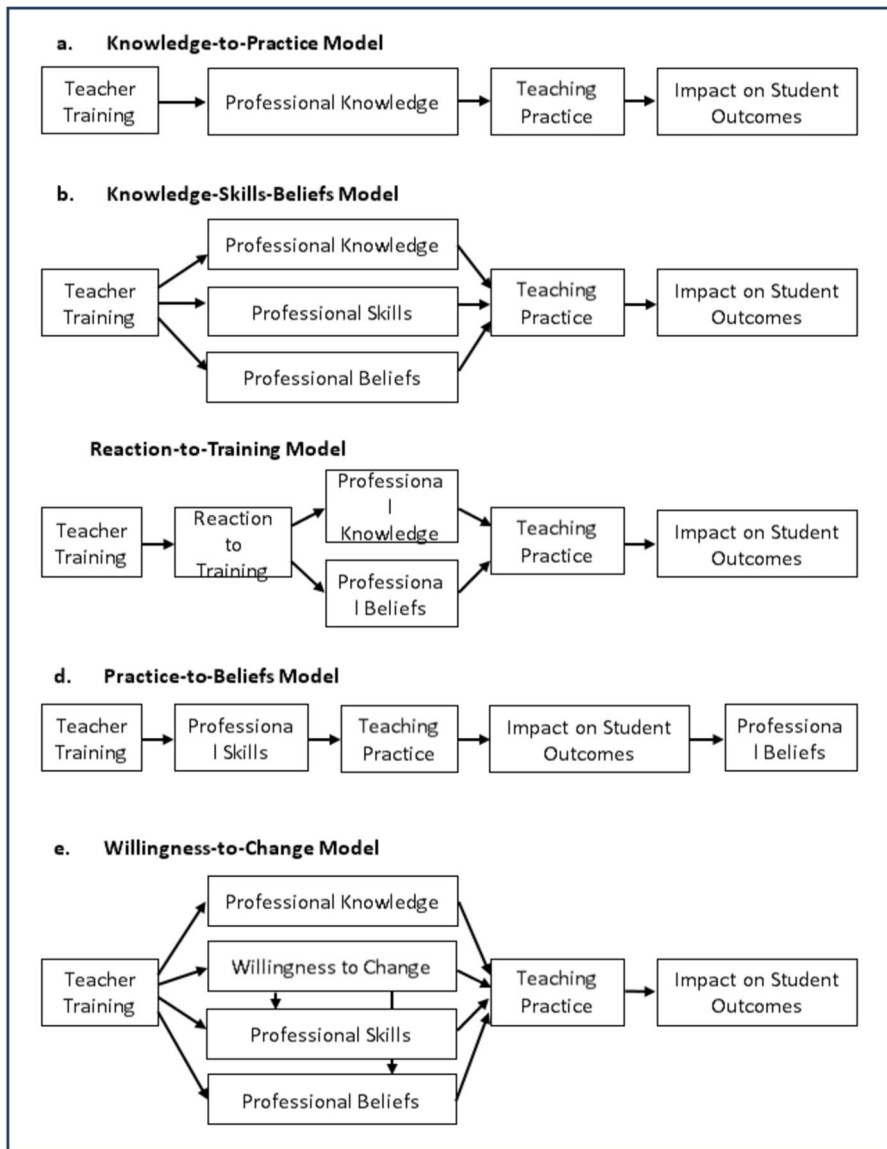
Fig. 1 Five models of learning transfer from teacher training to the classroom

Egert et al., 2018). However, they typically focus more on the structure and design of the training itself rather than on the underlying teacher learning processes, which are often described only briefly or remain largely unspecified.

Still, the assumptions of Knowledge-Skills-Beliefs models align with findings from studies showing that next to knowledge, training also impacts teachers' professional skills and beliefs regarding IE (e.g., Donath et al., 2023), and mere knowledge

about a new teaching approach is not enough to bring about changes in teaching practices (Cohen et al., 2020). Furthermore, it aligns with the theory of planned behavior (Ajzen, 1991), which posits that the perceived ability, combined with a positive attitude toward IE, influences subsequent behavior. However, it overlooks research focusing on teachers' professional beliefs, suggesting that slow and effortful processes are required to change the underlying beliefs that guide behavior (Lazdina & Daga-Krumina, 2023; Pajares, 1992). Empirical studies further debate whether correlations exist between changes in professional skills and beliefs and changes in teaching practices (Boeve-de Pauw et al., 2022; Boyle et al., 2020; Castillo et al., 2018; Wu, 2021).

Reaction-to-Training Model

Another type of transfer model commonly described and applied in training evaluation assumes that teachers' engagement in the learning process is mainly influenced by their perception and opinion of the training program (Kirkpatrick & Kirkpatrick, 2006; Lipowsky, 2010; Moore et al., 2009; Wade, 1985; see Fig. 1c and Table 1). This perspective suggests that the transfer process is shaped crucially by how participants perceive the training. While a pleasant atmosphere is often seen as key to engagement, meaningful learning and uptake of new approaches may arise from moments of discomfort. Recognizing the need for change, sometimes triggered by cognitive dissonance rather than comfort, can be a powerful catalyst for reflection. The literature remains inconclusive regarding the impact of training satisfaction on learning and transfer processes: while some studies report associations between satisfaction and self-efficacy and predicting changes in practice, although with small effect sizes (Rutherford et al., 2017), others find no predictive link between satisfaction and knowledge gain (Alliger et al., 1997; Goldschmidt & Phelps, 2010). When applying this kind of transfer model, researchers often limit themselves to the first steps of the model (Steele et al., 2016), investigating the training effect solely by focusing on teachers' self-reported satisfaction and self-perceived gains in professional knowledge, rather than observed behavioral outcomes (Cahapay, 2021).

Practice-to-Belief Model

Practice-to-Belief models (e.g., Cobb et al., 1990; Fogaça et al., 2024; Guskey, 2002) focus on the change processes in the classroom rather than what is learned during training. They assume that professional skills are acquired through training, enabling the implementation of reform practices, and observing positive impacts on students can lead to changes in teachers' professional beliefs (Guskey, 2002; see Fig. 1d and Table 1). Since belief change assessed directly after the training has seldom been a good predictor for actual change in teaching practice (Doherty, 2011; Strang, 2022), Practice-to-Belief models refrain from the assumption that first teachers' professional beliefs need to be changed in order to enable change in teaching practice (Ewing et al., 2018; Joyce, 1987). The key point is that it is not the training

itself, but rather the experience of successfully implementing new practices, that influences teachers' attitudes and beliefs (Guskey, 2002).

Several studies empirically support Practice-to-Belief models (Voelkel & Chrispeels, 2017; Zambak et al., 2017). For example, training programs could predict changes in teacher knowledge and practice, but not in attitudes (Strang, 2022). This aligns with studies showing that teachers' beliefs may be more challenging to change than their teaching practices (Behrmann & Souvignier, 2015), and that teachers who successfully and sustainably changed their teaching practices and professional beliefs often attributed this to positive experiences with their students (Bonner et al., 2019). This acknowledges that teachers' professional beliefs are generally stable and difficult to change (Lazdina & Daga-Krumina, 2023), and that these beliefs are shaped mainly by classroom experience, which in turn acts as an incentive for sustaining changes in practice (Bonner et al., 2019; Meiers & Ingvarson, 2005). However, assumptions based on Practice-to-Belief models focus more on long-lasting processes, including feedback loops, which can rarely be depicted in quantitative evaluation studies, as these typically assess the effects of training participation only once, shortly after training ends.

Willingness-to-Change Model

One component often mentioned as crucial for the learning and transfer process in teacher training is teachers' willingness to improve their current practice (Grohmann et al., 2014; Reinhold et al., 2018). Although closely connected to learning motivation, which is typically viewed as a prerequisite for the learning process and not as something that can be enhanced through teacher training (Scott et al., 2024), a growing number of researchers have proposed that teachers' willingness to change, as opposed to their learning motivation, should be included in transfer models for teacher training as a link between knowledge acquisition and implementation (Herawati et al., 2022; Osman & Warner, 2020; Scott et al., 2024; Wu, 2021). This aligns with general adult learning theories, which emphasize that the willingness to engage in change processes requires self-determination, subjective relevance, and an expectancy of achieving improvements (Eccles & Wigfield, 2000; Knowles, 1984). The failure to awaken participants' willingness to engage in change processes may also be one possible explanation for the implementation gap (Sims et al., 2023). Kennedy (2016) describes that this willingness can be activated through provocative questions that require deep reflection on one's behavior, creating 'aha' moments that can influence teachers' in-the-moment decisions. Models reflecting this call will be referred to as 'Willingness-to-Change models' (e.g., Azizatur et al., 2021; Gregoire, 2003; Kennedy, 2016; Lipowsky & Rzejak, 2015; Sims et al., 2023; see Fig. 1e and Table 1). These models assume that teacher training can enhance teachers' willingness to change and professional knowledge by provoking 'aha' moments and supporting changes in professional skills and beliefs, thereby triggering changes in teaching practices that can improve student outcomes. Support for this model can be found in recent studies, which show that readiness for change is a significant predictor of teachers' change in teaching practices (Azizatur et al., 2021) and attitudes

(Akin-Sabuncu & Calik, 2023). Accordingly, teacher trainings with compulsory participation, a practice that reduces or even ignores teachers' motivation to engage in change processes, have been reported to have lower effect sizes (Donath et al., 2023; Kennedy, 2016).

Research question

To ensure that the legal mandate to implement IE translates into effective classroom practice, teachers must be adequately supported in navigating the complex demands of inclusive teaching. Teacher training is crucial for providing this support cost-effectively to teachers. To investigate its effectiveness, it is essential to be aware of the relevant steps in the transfer process. Education research has established strong theories on student learning. However, a lack of a well-developed understanding of how teachers learn and how they can effectively integrate new ideas into their teaching practices remains (Kennedy, 2016). Currently, there is a consensus among models on learning transfer that teachers need to be informed about school reforms, such as IE, in order to change their current practice and enable improvements at the student level (Boylan et al., 2018; Nīmante et al., 2025). In addition to this straightforward model, numerous models incorporate additional elements and describe them as relevant, differing in the number of elements described, the number of steps, and the sequence in which the elements interact with one another.

A discussion about the existing theoretical foundation of teacher training is needed to improve and systematize research in this field (Nīmante et al., 2025). However, no empirical synthesis of transfer theories has been conducted (Blume et al., 2010). To design training evaluations effectively, it is essential to investigate whether more detailed descriptions of learning processes provide added value and to identify which indicators are most predictive of the transfer process. Particularly in the context of IE, as there is no clear-cut answer to what successful implementation looks like, it is essential to support teachers effectively in developing a deep understanding of concepts, allowing them to adapt them to the specific conditions of their classrooms and schools. To address this research gap, the present study aims to test whether the described models differ in their predictive value for data on teacher training for IE.

The present study used secondary data to apply meta-analytic structural equation modeling (MASEM). MASEM refers to the application of structural equation models (SEM) to meta-analytic data using correlation matrices, enabling the investigation of complex hypotheses through secondary data analysis (Bergh et al., 2016). This method allows for combining results from multiple primary studies to increase statistical power and generalizability and further facilitate testing complex theoretical models that are challenging to assess in individual studies. Investigating processes of learning transfer in the context of teacher training through meta-analysis is also more general than examining one specific training program alone. Since teacher trainings can be designed in very different ways, meta-analysis can further investigate whether features of the training design influence the transfer process. MASEM further allows testing the explanatory value of different theoretical models

by comparing competing models (Bergh et al., 2016). This is what was done in the present study, which investigated which of the five models of learning in teacher training (Fig. 1) had the best fit on the available accumulated data.

Methods

The systematic literature search, screening, and selection of studies, as well as the data collection, followed the PRISMA guidelines (Salameh et al., 2020).

Literature Search

The data collection process was based on a previous study (Donath et al., 2023). The systematic literature search used the same databases and search terms, but with different study inclusion criteria. The databases PsycINFO, Web of Knowledge, and ProQuest (ERIC, Education Research Complete, Dissertation and Thesis) were searched 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 assistant OR educators AND school OR K-12 OR kindergarten OR preschool OR vocational college (for more detailed information see Supplement C). To reduce the risk of publication bias, unpublished studies were included intentionally. The first literature search was conducted in April 2021 and updated in July 2024.

Dependent Variables

For the above-described theoretical models (Fig. 1), seven relevant dependent variables were identified. These variables had in common that teacher training should influence them in some way for a successful transfer process. Since the operationalization of these variables could vary, especially in field studies, eligibility criteria for the measures of all these variables were set (for additional information, see Supplement C). To be included in the meta-analysis, (1) teachers' knowledge needed to be assessed via knowledge tests or self-reports. These assessments typically focus on contextual knowledge related to IE (e.g., legal framework, disability diagnostics, and concept of IE). While it has been debated whether self-assessed knowledge reflects self-efficacy (Kadluba et al., 2024), this concern mainly applies to self-ratings of specific teaching methods. In the context of IE, knowledge assessments differ from evaluations of one's ability to implement inclusive practices. Moreover, a recent meta-analysis on teacher training for IE found no significant differences between tested and self-rated knowledge, and therefore combined these variables into a single variable (Donath et al., 2023). Regarding teachers' skills, theoretical models differentiate between acquiring skills and applying them in practice (e.g., Desimone, 2009). For the present study, (2) teachers' skills were defined as whether teachers feel capable of using these skills, along with the theoretical descriptions

that skills can be acquired in the actual training; hence, we included assessments of self-efficacy for IE and perceived competence to implement inclusive practices. (3) Teaching practices reflect the application of these skills. Eligible measures of teaching practices assessed either the implementation quality or the use of inclusive teaching methods via self-reports, in-person observations (time sampling or quality ratings), or student ratings. (4) Teachers' beliefs regarding IE (general attitudes towards IE and perceptions of inclusive teaching methods), (5) willingness to change (motivation to engage in change processes towards IE), and (6) satisfaction with the training were always based on self-reports. Measures on the (7) student level could reflect either the students' achievement or behavior. It was chosen to include differing measures because the content of teacher training can target various aspects at the student level, and to include as much data as possible at this rarely assessed level. Student achievement needed to be assessed using standardized tests, performance in tasks given by the teacher, or meeting the individual education plan goals. Assessments of students' behavior needed to be based on students' self-reports, teacher reports, or observational measures assessing on-task behavior and engagement in classroom tasks, attendance in school, challenging behaviors, and supportive behavior towards other students.

Inclusion Criteria

Studies had to fulfill the following criteria to be included in the analysis: Studies had to (1) report data on teacher training for IE of (2) teachers and/or their students with (3) quantitative measures, (4) either in cross-sectional designs, longitudinal within-group (pre vs. post) change designs, or in comparisons with a control group. Studies had to (5) provide data on at least two of the following variables: teacher knowledge, skills for IE, teaching practices, beliefs regarding IE, willingness to change, satisfaction with the program, and student achievement and learning behavior.

Studies were excluded when the teacher training was solely based on coaching (i.e., one-on-one training of teachers). Studies comparing special education with general education teachers were excluded because the present study was interested in teacher training on IE, not in comparing the two main professions involved in implementing IE in schools. When results were reported in more than one publication, the publication with the most detailed information was included in the analysis (e.g., Kuok et al., 2022; Yada et al., 2018). Furthermore, if data from one study were split across several publications, and it was clear that the data originated from the same study, the data from the various publications were summarized and treated as a single study in the analysis (see Supplement E).

To ensure the reliability of the screening process, two raters independently screened a sample of 500 abstracts. Disagreements regarding the inclusion or exclusion of studies were resolved through discussion and consensus. Inter-rater reliability was assessed using Krippendorff's alpha (2011), indicating a high agreement between the raters ($\alpha = 0.907$).

Coding

For a transparent coding process, a condensed version of the coding handbook by Donath et al. (2023) was used (see Supplement A). The publications selected during the screening process were subsequently coded according to various characteristics, including the source of information (publication type and country), study design (sample size, sampling method, allocation to conditions, attrition, and conduct with missing data), participant characteristics (profession, experience with IE, and school type), the design of the teacher training (duration, topics covered, practice opportunities, certification received, and participation with colleagues), obtained results (data collection method, instrument, and analysis approach), and the intercorrelations of various outcome measures (teachers' knowledge, self-efficacy, teaching practices, beliefs, willingness to change, satisfaction with teacher training, and student achievement and behavior).

When multiple samples were reported in a single study, all samples that fulfilled the inclusion criteria were included. Samples were used as the unit of analysis in the present study, as the effect size metric, Pearson's *r*, was used to quantify the associations between training, teachers' knowledge, self-efficacy, teaching practice, professional beliefs, willingness to change, satisfaction with the training, and students' achievement and behavior. When correlations were not reported, but other data were available on relations between variables of interest, the R package *esc* (Lüdecke, 2019) was used to calculate the effect size in the required metric. For publications after 2010, authors were contacted when data of interest were not reported. In sum, 76 authors were contacted; 18 responded, and 14 provided the missing data.

To evaluate the risk of bias in individual studies, we modified the Medical Education Research Study Quality Instrument (MERSQI; Reed et al., 2007) for use in the present meta-analysis (see Supplement B). The MERSQI was chosen because of its reported high inter-rater reliability. Our adapted version evaluated the following aspects: (1) study design, (2) response rate, (3) sampling method, (4) allocation to conditions, (5) use of standardized instruments, (6) internal structure of the instrument, and (7) reporting of missing data handling, with a maximum of 9.5 points achievable (see Supplement C). The study design was rated as follows: 1 point for cross-sectional designs, 1.5 points for pre/post change evaluation with an intervention group only, 2 points for post-intervention comparisons with both an intervention and control group, and 3 points for studies with pre/post change evaluation in both the intervention and control groups. The response rate was rated as follows: zero points when no information was given, 0.5 for a response rate below 50%, one point for a response rate between 50 and 75%, and 1.5 points for a response rate above 75%. The remaining aspects were rated with either one or zero points based on the criterion of whether the random sampling procedure, random allocation to conditions, the use of standardized instruments, the reported internal structure of the evaluation instrument, and the handling of missing data were provided or not.

Analysis

First, descriptive statistics regarding the number of extracted correlations and the characteristics of included studies were created. Second, potential moderators (study quality, study design, and participants) were analyzed to assess their influence on the reported effect sizes using meta-regression. Further, we analyzed the potential influence of publication bias. This is described in more detail below. Third, MASEM was used to test the theoretical models (Fig. 1), applying random-effects one-stage MASEM (Jak & Cheung, 2020). It offers a streamlined and integrated approach to simultaneously synthesize effect sizes and test structural models, making it particularly useful for complex datasets or heterogeneous studies. Random-effects modeling was chosen to account for the expected between-study heterogeneity resulting from the variety in sample characteristics and training designs among the included studies.

Furthermore, one-stage MASEM can handle missing data in the correlation matrices using full information maximum likelihood (FIML), which is the case in the currently included set of studies. One-stage MASEM is applied to the pooled correlation matrix of all correlations reported in the primary studies. A separate pooled correlation matrix was therefore calculated for each theoretical model (Fig. 1), as studies differed in the number of reported correlations and could not be used for each model, but only for models for which they provided at least one of the required associations (see Supplement F). To enhance the comparability of model calculations, the analyses are repeated using a dataset that allows for the computation of all models. Additionally, as study design and study quality can also influence effect sizes, the analyses are repeated using only data from studies with randomized or quasi-experimental assignment to conditions (RCTs and quasi-experimental controlled trials). Model fit was evaluated as acceptable or good when the CFI and TLI were higher than 0.9 and 0.95, respectively, the RMSEA was smaller than 0.06, and the SRMR was smaller than 0.1 (Hu & Bentler, 1999). The MASEM analyses were conducted using WebMASEM (Jak et al., 2021). In the path models, standardized path coefficients are reported. Effect sizes of $r = 0.1$ were interpreted as small effects, $r = 0.3$ as medium effects, and $r = 0.5$ as large effects (Cohen, 1988).

Publication Bias

Several methods were applied to evaluate the risk of publication bias, as suggested in the literature (Rothstein et al., 2005; Siegel et al., 2021) and since there is no well-established way to detect publication bias in MASEM (Cheung, 2014). To reduce publication bias, unpublished studies were intentionally included in the present study, and moderator analyses were conducted to investigate potential differences between published and unpublished studies. Although correlational meta-analysis is relatively robust against the effects of publication bias (Chow & Ekholm, 2018), Egger's regression tests and funnel plot analyses were conducted with the data from published studies to assess potential bias. For these analyses, dependencies among

effect sizes from one sample were accounted for using multilevel models to calculate Egger's regression test (Fernández-Castilla et al., 2021) and create the funnel plots. In Egger's regression test, the standard errors of the effect sizes served as the predictor variable, and the tests examined the relationship between the effect sizes from each study and their corresponding standard errors to detect small-study effects. Power-enhanced funnel plots (Kossmeier et al., 2019) were used to inspect potential asymmetry in the effect size distribution visually and to obtain information on the power of individual studies to detect the presence of effects.

Results

Description of Included Studies

The literature search yielded 14,226 records after removing duplicates, and 228 publications were identified as fulfilling the inclusion criteria during the screening process (see Fig. 2 for more detailed information; Supplement D and K provide further details). These publications reported data on 244 samples and provided 1084 effect sizes.

Then, 62% ($k=142$) of the included studies were journal articles, and the remaining studies were unpublished dissertations ($k=72$), project reports ($k=7$), conference proceedings ($k=6$), and preprints ($k=1$). The majority of studies were conducted in North America ($k=109$; 47.6%), followed by Asia ($k=48$) and Europe ($k=39$). In total, 61,766 teachers participated in the primary studies, and 55,843 students, of whom 21% had a diversity characteristic linked to potential educational disadvantage. Of the participating teachers, 74.7% were general education teachers,
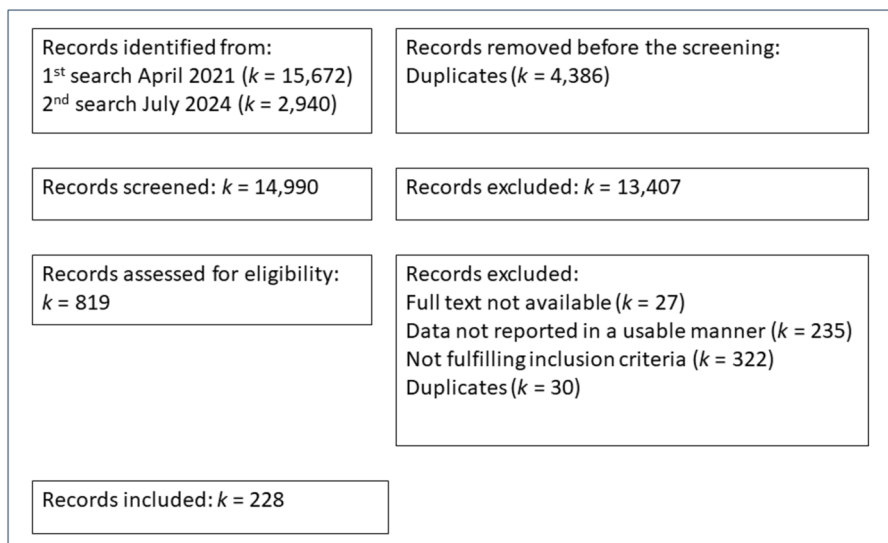
Records identified from: 1st search April 2021 (k = 15,672) 2nd search July 2024 (k = 2,940)	Records removed before the screening: Duplicates (k = 4,386)
Records screened: k = 14,990	Records excluded: k = 13,407
Records assessed for eligibility: k = 819	Records excluded: Full text not available (k = 27) Data not reported in a usable manner (k = 235) Not fulfilling inclusion criteria (k = 322) Duplicates (k = 30)
Records included: k = 228	

Fig. 2 PRISMA flowchart representing study selection

8% were special education teachers, 6% were administrators, 1.4% were teaching assistants, and 9% were pre-service teachers. Then, 77% of the participating teachers had experience with IE, and 74% were teaching in inclusive classrooms.

Twenty-four studies were conducted in kindergarten, 71 in primary school, and 45 in secondary school. The remaining 88 studies did not indicate information regarding the school type. Most teacher trainings focused on special educational needs (SEN; $k=189$); among these, 51 specifically addressed one type of SEN. Twenty-seven studies focused on other diversity features, primarily at-risk students ($k=6$), second language learners ($k=6$), and LGBTQ individuals ($k=5$). In comparison, 13 studies addressed the presence of multiple diversity features simultaneously.

Most training programs ($k=85$) spanned several weeks, with practice phases conducted in their classrooms between training sessions. Trainings lasted between 2 h and 4 school years (median$=12$ weeks), and post-data were assessed on average 3 months after the first training session, ranging from assessments directly after the training ($k=73$) to 6 years after the first training session.

Study Quality and Publication Bias

The study quality among the included studies was generally low ($M=4.63$, SD$=1.28$, range$=2$–9.5). The study with the highest rated study quality applied a controlled randomized trial design with pre- and post-assessments in both the intervention and control groups, random sampling and allocation to conditions, used standardized instruments, and reported on the internal structure of these scales and how missing data were handled (Yoshikawa et al., 2015), while the study with the lowest study quality applied a cross-sectional design with a convenience sample and reported no information on response rate, applied instruments, or how missing

Table 2 Descriptive statistics of study quality rating

Characteristics	Number of studies	%
Study design	112	49
Cross-sectional	71	31
Pre/post-change in intervention group only	19	8
Intervention/control group comparison only after training	27	12
Pre/post-change with intervention and control group		
Response rate	142	62
Not reported or not applicable	72	32
$<50\%$	7	3
50–75%	6	3
$>75\%$		
Random sampling procedure	23	10
Random and quasi-experimental allocation to conditions	54	24
Use of standardized instruments	71	32
Internal structure of instrument reported	143	64
Handling of missing data reported	86	38

data were handled (Bergren, 1997; Gaastra et al., 2020). The distribution among the aspects used to evaluate study quality is presented in Table 2.

Publication status was not associated with the reported effect sizes ($F[1, 1041] = 0.67$, $p = 0.41$). However, study quality was associated with the reported effect sizes ($F[1, 1041] = 6.54$, $p = 0.01$), with lower study quality being associated with larger effect sizes ($\beta = -0.03$, see Fig. 3). Egger's regression test among the published studies was also significant ($F[1, 681] = 118.28$, $p < 0.001$), and a clear asymmetry was evident in the funnel plot (see Fig. 3). However, the direction of this observed asymmetry was *contrary* to small-study effects, as studies with *larger* samples reported *larger* effect sizes. Further analyses showed that this might have been because most studies with higher power also investigated more extended training programs. Accordingly, smaller standard errors of effect sizes were associated with both larger effect sizes ($r = -0.14$ [-0.22; -0.05], $p = 0.002$) and longer training durations in the studies ($r = -0.19$ [-0.27; -0.10], $p < 0.001$). The median power of studies appeared to be generally low (22.3%). However, this estimate was based on the summary estimate of all correlations overall. As bivariate associations varied across all pairs of variables (see Table 3), this estimate should not be considered accurate for the associations between all variables in the tested models.

Moderation Analyses

Intervention studies reported larger effect sizes (overall $r = 0.35$, 95% CI [0.31; 0.39]) than cross-sectional studies (overall $r = 0.24$ [0.21; 0.27]; $F[1, 1041] = 22.36$, $p < 0.001$). Among the intervention studies, study design (cross-sectional, within-group pre vs. post, and control group) was not associated with the reported effect sizes ($F[2, 510] = 0.41$, $p = 0.66$). Publication year, the continent where the study was conducted, and school type did not significantly influence the reported effect sizes (all $F < 3.4$, all $p > 0.06$). The educational background of the teachers had no effect (all $F < 0.6$, all $p > 0.4$). However, studies that included teaching assistants reported larger effects (overall $r = 0.39$ [0.21; 0.57]) than those that involved only teachers (overall $r = 0.29$ [0.26; 0.32]), and the more teaching assistants were included, the larger the reported effect sizes ($F[1, 1041] = 7.05$, $p = 0.008$, $\beta = 0.22$). Whether teachers were teaching inclusive classrooms or having inclusive teaching experience did not show significant influence (all $Fs < 1.1$, all $ps > 0.35$). For additional analyses regarding possible moderators, see Supplement J.

Testing Transfer Models

We created a pooled correlation matrix using the correlation coefficients from the collected studies to test the five models of transfer processes. The pooled correlation matrix is presented in Table 3, along with the sample size (i.e., the number of samples) for each correlation. However, the analysis of each model was based on separate correlation matrices, which included only the variables specific to that model.

The fit indices of the different models are shown in Table 4 (for additional information, see Supplement F). CFI was below 0.9 for all models except for the

Table 3 Meta-analytic pooled correlation matrix

	1	2	3	4	5	6	7	8
(1) Training	x	81	160	102	174	46	17	52
(2) Knowledge	.371	x	41	26	44	13	5	6
(3) Skills	.265	.241	x	31	80	22	9	6
(4) Teaching Practice	.269	.180	.284	x	33	13	7	17
(5) Beliefs	.183	.147	.282	.199	x	22	10	12
(6) Willingness to Change	.199	.059	.318	.391	.350	x	5	7
(7) Satisfaction with Training	.216	.288	.310	.235	.425	.368	x	4
(8) Student Behavior	.236	.218	.284	.249	.315	.197	.415	x

Pooled correlation coefficients are shown in the lower triangle, and sample size (number of samples) is shown in the upper triangle. Colors highlight the magnitude of the correlations, with orange indicating the strongest associations and white indicating the weakest.

Willingness-to-Change model, which achieved an acceptable model fit. None of the models met the 0.9 threshold for the TLI, which can be interpreted as a variant of the CFI that considers model parsimony. Low values in this index may have been due, in part, to the relatively small size of the models compared to their large number of parameters. However, the Willingness-to-Change model came closest to the threshold with a value of 0.808. RMSEA was below the cut-off value of.06 for all models, indicating a good model fit. This indicator again suggests that the Willingness-to-Change model is the best fit.

Regarding the SRMR, only the Willingness-to-Change model achieved a value below the cut-off of 0.1. The fit indices indicated that the Willingness-to-Change model had the best fit in comparison to the other models (three out of four indicators meeting the criteria for an acceptable or even good model fit). However, it is worth noting that some path coefficients were based on small databases (Table 3), particularly those paths involving teachers' satisfaction with the training and student outcomes.

Since the fit indices were based on (slightly) differing data sets, which may distort direct model comparisons, the analyses were repeated, using strictly the same data for all models ($k = 163$, Supplement G). The results with these data yielded lower fit indices overall (see Supplement I); however, the Willingness-to-Change model still had the best fit. Two indices indicated a good fit, while two fell slightly below the cut-off value for an acceptable fit; however, all indices had higher values than in the other models. In the analysis using only data from studies with randomized or quasi-experimental assignment to conditions (RCTs and quasi-experimental controlled trials, $k = 58$), the Willingness-to-Change model also demonstrated the best fit. However, the values were again lower compared to the main analysis (see Supplement H). Three out of four fit indices indicated a good fit, while one fell slightly below the cut-off value for an acceptable fit. Based on this analysis, a second model, the Knowledge-Skills-Beliefs model, also showed a good fit in three out of the four indices, with only slightly lower values than the Willingness-to-Change model.

In summary, based on the three analyses (main analysis, analyses using the same database for all models, and the RCTs only), the Willingness-to-Change model had
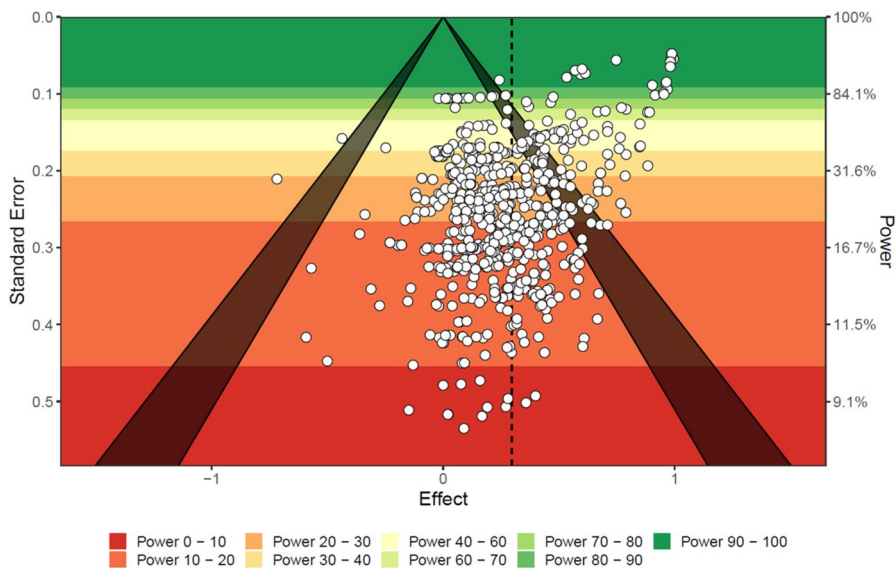
Fig. 3 Power-enhanced funnel plot

Table 4 Goodness-of-fit statistics for the five models (main analysis)

Model	CFI	TLI	RMSEA	SRMR
Knowledge-to-Practice Model	0.749	0.248	0.015	0.166
Knowledge-Skills-Beliefs Model	0.859	0.698	0.011	0.114
Reaction-to-Training Model	0.722	0.479	0.011	0.201
Practice-to-Belief Model	0.812	0.647	0.012	0.197
Willingness-to-Change Model	0.918	0.808	0.008	0.099

the best fit with the available empirical data. The associations between the variables in the Willingness-to-Change model are shown in Fig. 4. Participation in teacher training had a large effect on teachers' professional knowledge ($\beta=0.81$, 95% CI [0.64, 0.97]) and moderate effects on teachers' willingness to change ($\beta=0.37$, 95% [0.23, 0.51]), self-perceptions of professional skills ($\beta=0.43$, 95% [0.28, 0.57]), and a small to moderate effect on teachers' professional beliefs ($\beta=0.25$, 95% [0.15, 0.36]). Teachers' willingness to change positively and significantly influenced perceptions of their professional skills ($\beta=0.34$, 95% [0.24, 0.45]) and their professional beliefs ($\beta=0.38$, 95% [0.30, 0.46]), both with moderate effect sizes. Changes in teaching practices were influenced with small to moderate effect sizes by teachers' knowledge ($\beta=0.30$, 95% [0.15, 0.44]) and willingness to change ($\beta=0.26$, 95% [0.06,.046]) and with a small effect size by teachers' perception of their skills ($\beta=0.12$, 95% [0.001, 0.25]). Teachers' professional beliefs did not significantly influence their teaching practice ($\beta=0.08$, 95% [−0.09, 0.25]). Changes in teaching
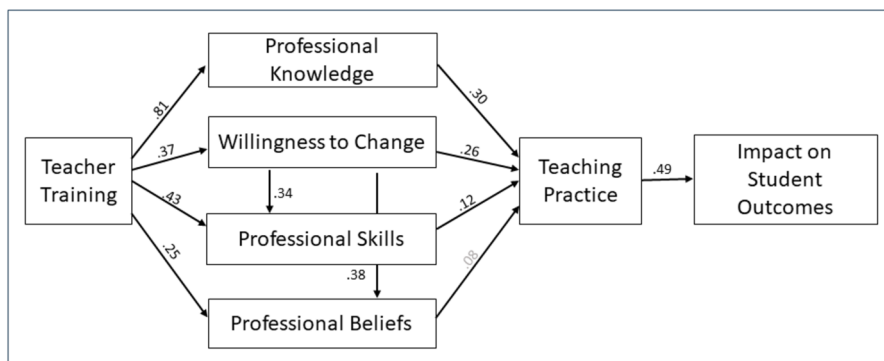
Fig. 4 Path results of the Willingness-to-Change model

practice had a moderate, significant effect on student outcomes ($\beta = 0.49$, 95% [0.26, 0.72]).

The patterns proposed by this model are also reflected in the base correlations, showing minimal associations between knowledge and willingness to change, small associations between knowledge and skills and beliefs, but stronger associations between willingness to change and skills, beliefs, and teaching practice. An overview of the path estimates among the three analyses for all models is presented in Supplement I.

Discussion

This study aimed to compare models that describe transfer processes in the context of teacher training on IE regarding their fit to the currently available empirical data. For this purpose, data from primary studies were collected, examining teachers' participation in training programs on IE concerning seven variables described in the theories as relevant to the transfer process. IE was chosen as the topic for the teacher training programs as it aims to improve learning conditions for all students and serves as a useful first step in examining the transfer processes of teacher training. In addition, IE presents teachers with the challenge of navigating complex learning processes in their daily professional lives, as it requires fundamental changes in their approach to teaching. Since teacher training is the most common method for supporting in-service teachers in this change process, gaining insight into the transfer process is crucial. The present study tested models that describe this transfer process using secondary data analysis.

The models were calculated three times: first, based on all available data, resulting in different sample sizes; second, based on studies providing data for all models; and third, based on studies with random and quasi-experimental designs. Based on these analyses, the Knowledge-to-Practice model, the Reaction-to-Training model, and the Practice-to-Belief model all appeared to be not well-suited to explain the available empirical data on the transfer process. The analyses suggested that

knowledge acquisition during training alone cannot predict transfer processes into the classroom and ultimately to students, as described in the Knowledge-to-Practice models. Furthermore, the analyses indicated that the evaluation of the training itself may not be the sole determining factor for subsequent learning and transfer processes, as proposed by the Reaction-to-Training models, since these models also failed to satisfactorily explain the empirical data. Practice-to-Belief models assume that shifts in professional beliefs occur gradually and require significant effort and reflection, primarily resulting from positive experiences with changed behavior. This assumption was not supported by the analyses either, possibly because post-measurements were conducted too shortly after the training to capture such delayed changes.

Regarding the Knowledge-Skills-Beliefs model, the results differed based on the data the model was fit to. Based on the RCT and quasi-experimental studies only, the model fit was reasonable, meeting three criteria, one of which was close to the cut-off value. However, the Knowledge-Skills-Beliefs model, which assumes that, in addition to acquiring knowledge, participants develop skills and more positive beliefs toward the training content, failed to achieve an acceptable model fit to describe the data adequately, based on the full available data and the studies that provided data for all models.

However, one group of models, the Willingness-to-Change models, achieved a good model fit in three out of four indices and, therefore, represented a framework well suited to describing the available data on the transfer processes. However, attention should be paid to the factors distinguishing the suitable models from the less suitable ones. Models summarized in the present study, under the term Willingness-to-Change model, differed from the other models primarily in their assumption that teacher training programs can influence the willingness to make changes in practice and that this willingness can drive changes in professional beliefs and teaching practices.

The current analysis supported this assumption as teacher training showed a significant influence on changes in teachers' willingness to engage in the change process, and this further resulted in significant influences on changes in teachers' skills, beliefs, and teaching practices, a pattern that also shows in the base correlations, and is in line with previous research (Akin-Sabuncu & Calik, 2023; Azizatur et al., 2021). While all relevant paths were statistically significant across different model variants, the size of the path estimates varied depending on the dataset used for model estimation, likely due to the limited amount of data available for associations involving willingness to change in the reduced datasets.

If the training enhances participants' willingness to actively attempt changes in their teaching, it increases the likelihood that these changes will also have a positive impact on students. Multiple reviews have suggested that teachers' motivation to participate in professional development, such as teacher training, should be strengthened (Dunst et al., 2015; Kennedy, 2016; Timperley et al., 2008). However, it may be that the motivation to implement concrete changes triggered during the training itself is the key factor in achieving long-term improvements for students.

However, our analyses did not support all the assumptions described in the Willingness-to-Change model. Empirically, teachers' professional beliefs were

influenced by participation in teacher training and teachers' willingness to change; however, these beliefs did not predict realized teaching practices, a pattern that was consistent across all three analyses. This raises questions about why only some cascading effects occur and calls for future research to investigate the relationships between the variables strengthened through training. Interestingly, the same assumption was made in the Knowledge-Skills-Beliefs model, which was also not supported by the present results. Notably, even in the RCT subsample, where this model achieved good overall fit, the path from beliefs to teaching practice remained non-significant (see Supplement H and I). This finding further contradicts the prominence of attitude-focused research in the IE literature, which is based on the Theory of Planned Behavior (Ajzen, 1991), assuming that internal beliefs drive teaching behavior. One possible explanation is that the beliefs captured in evaluation studies reflect surface-level beliefs that can be influenced in the short term by teacher training and social desirability but may have a limited impact on teaching behavior, especially under real-world school conditions. Moreover, the constructs beliefs and practice themselves are conceptually distant as beliefs are typically assessed through self-reports in questionnaires, which may reflect idealized self-perceptions, whereas teaching practices (although also often self-reported) refer more directly to concrete actions and actual implementation in the classroom. In this context, potential differences between general attitudes towards IE and the perception of inclusive teaching strategies (both of which were included under the construct of professional beliefs in the present study) should be acknowledged. It is conceivable that changes in the perception of inclusive strategies are more closely linked to changes in teaching practice. However, due to limited data availability, this aspect could not be examined separately in the current analysis. Further, the path from professional beliefs to teaching practice could be partially attributable to underpowered primary studies or inconsistent measurement strategies, which will be discussed below. Still, the influence of attitudes on behavior has been questioned before (Gülsün et al., 2023). The present findings thus align with previous research, which shows that attitudes towards IE have a relatively minor influence on teaching behaviors (Gülsün et al., 2023; Sharma & Mannan, 2015). Additionally, the research indicates that teaching behaviors are more likely to change than professional beliefs through training participation (Behrmann & Souvignier, 2015). Although differently proposed by both the Willingness-to-Change model and the Knowledge-Skills-Beliefs model, belief change triggered during the teacher training and assessed directly after the training ended may not yet be strong enough to influence teaching practice in the classrooms but may take more time to unfold and support long-term changes in teaching practices (Guskey, 2002).

Still, attitudes and changes in attitudes are most commonly used to measure the success of teacher training participation in improving student outcomes (Donath et al., 2023). Instead, self-efficacy should be assessed, as it aligns more closely with behavior and exhibits greater predictive value (Gülsün et al., 2023). Alongside self-efficacy, willingness to change should be more consistently assessed in evaluation studies. Motivational learning theories (Eccles & Wigfield, 2000; Knowles, 1984) emphasize the importance of relevance and self-determination in the change process—factors that are closely reflected in the construct of willingness to change.

Assessing this psychological readiness to alter teaching practices may therefore be a stronger predictor of actual change than professional beliefs. Although the beliefs regarding IE in the present study did not significantly influence changes in teaching behavior, it is still conceivable that teachers' attitudes affect how they interact with students, as well as their teaching methods. The results of this study contradict the literature on IE, which repeatedly argues that attitude is crucial for the implementation of IE (Lindner et al., 2023). Therefore, the significance of attitude for teaching practice, and in particular its effects on students, should be further examined in detail in future studies.

In general, the proposed factors influencing teaching practice, knowledge, willingness, skills, and beliefs did not show large effect sizes on changes in teaching practices, and 73% of the observed variance of changes in teaching practices remained unexplained. A possible explanation could be that what is realized in the classroom is not only influenced by the proposed variables but also by the circumstances in the school, such as a supportive administration and colleagues, as well as available resources and materials, which do not change simply because one teacher participates in a teacher training program. The present study provides some hints that support in schools can enhance change processes, as training conducted for teaching assistants and teachers together led to larger effect sizes. Several factors could explain this pattern. First, teaching assistants often have less formal training than teachers, which may make them more receptive to new learning. Unlike experienced teachers, they may not need to reconcile new knowledge with previously acquired professional beliefs and practices, allowing for a more straightforward application of training content. Second, teaching assistants may enhance collaboration between teachers and support staff, leading to more effective implementation of new practices. Future studies that utilize and examine the Willingness-to-Change model could, for example, incorporate school framework conditions into their analyses. Although these conditions, along with teachers' learning prerequisites, were not considered in the present analysis, they should not be overlooked. Since actions never take place in a vacuum, and school environments can vary significantly while also having a substantial impact on implementation (Nīmante et al., 2025), these factors should be investigated. They may also contribute to explaining the variance in the effectiveness of training programs.

Beyond contextual factors related to the school environment, future research should also explore how willingness to change can be actively supported and fostered during teacher training, as well as how it is related to other aspects addressed in teacher training, such as the acquisition of knowledge, skills, and beliefs. Training success depends mainly on teachers' ability to assess what is valuable for their classrooms and what is not (Nīmante et al., 2025). This ability should be actively fostered through training design by encouraging reflection on current practices, evaluating them in light of new approaches, and identifying meaningful changes. These can be triggered by reflective processes, for example, by encouraging teachers to reflect on specific students or teaching challenges and to relate the training content directly to their own classroom experience and prior knowledge. Adult learning and motivation theories (Eccles & Wigfield, 2000) suggest that perceiving content as relevant and practically feasible is essential for engagement and change. To support this, teachers need dedicated time

and space to plan implementation, including determining first steps, adapting methods to their specific contexts, and developing concrete integration strategies. Kennedy (2016) found that training programs focused solely on knowledge delivery tended to produce smaller effect sizes than those designed to spark insight and motivation. If motivation is a key driver of change, it is worth reconsidering the role of mandatory professional development, as simply requiring participation does not guarantee meaningful learning (Kennedy, 2016). Instead, whole-school training should prioritize fostering motivation, emphasizing the benefits of new practices, and actively engaging educators in the change process. To evaluate the effectiveness of such measures, willingness to change should be assessed before, after, and well beyond the training to capture both immediate and long-term effects.

In summary, these results suggest that professional development may have a positive impact on both teachers and students. Since the most complex model had the best fit, it can also be concluded that the learning and transfer processes are accordingly rather complex as well. The present results also support a recent research trend that emphasizes the importance of teachers' willingness to change, suggesting that this motivation should be strengthened in training programs. Without a focus on willingness to change, teacher training may increase teacher knowledge but fail to produce concrete changes in classroom practice.

Limitations

A key limitation of this meta-analysis is rooted in the considerable variability of the included studies, as data were drawn from rather diverse contexts. While the applied MASEM approach allowed for synthesizing findings across studies, it cannot replace direct hypothesis testing using primary data. This is particularly relevant for the Willingness-to-Change model and other tested models, which should ideally be evaluated using datasets specifically designed to capture all relevant variables. However, no single study among the collected studies included all the variables necessary to test the Willingness-to-Change model directly, nor was a complete dataset available for most of the other models. This also influenced the number of samples on which the models were based. To address this issue, the analyses were repeated using a common data subset that included all variables, which yielded similar conclusions, albeit with lower fit indices. Moreover, several model paths were based on very limited data, particularly those involving teachers' reactions to training and student-level outcomes. Although assessments at the student level are often complex to plan (Kowalski et al., 2020), they are highly relevant to evaluating the success of teacher training, as this is the level where improvements are targeted.

Another limitation arises from the heterogeneity of measurement methods used across the primary studies. Variables were assessed using a wide range of instruments (see Supplement C and E), which may capture different facets of the same construct and vary in their alignment with the specific training content. This heterogeneity is partly due to the broad spectrum of professional development programs included in the review, which targeted diverse aspects, particularly concerning teaching practices and student outcomes, and thus employed different tools to capture

change in those areas. In addition, these constructs are often associated with higher assessment effort (e.g., classroom observations or student testing), making them less frequently assessed in primary studies. Many measures relied on self-report, although constructs like knowledge or teaching practice could also be assessed through objective tools such as tests or classroom observations. A recent meta-analysis (Donath et al., 2023) found no systematic differences in training effects based on measurement method, including teachers' knowledge and teaching practice, as well as student outcomes. In the present meta-analysis, such effects were difficult to isolate, as effect sizes represent associations between variables and may therefore be shaped by the combination of methods used. Importantly, measurement methods can also influence the similarity between constructs; for example, several self-reported measures may appear more strongly associated with each other than when one of them is assessed through an objective method. Some constructs inherently rely on self-reported measures (such as beliefs or self-perceived skills), which contributes to their methodological similarity; however, the previous meta-analysis demonstrated that these constructs still differ significantly from one another, indicating that they capture distinct underlying dimensions. Due to data limitations, it was not possible to restrict analyses to datasets using only one method per construct. Therefore, we were dependent on the available heterogeneous data, underlining the need for future studies with more consistent and qualitative measurement approaches.

Thirdly, effect sizes reported in the included studies may be overestimated. This risk stems from two main factors: the generally low statistical power and quality of many primary studies, as well as the frequent use of post-measurements conducted immediately after the intervention (Donath et al., 2023; Kennedy, 2016). Both factors are known to inflate effect estimates, a pattern also observed in the present study, where lower study quality was associated with larger effect sizes, particularly as the impact of training often fades over time (Sims et al., 2023). While some argue that the effects of teacher training may develop gradually (e.g., Clarke & Hollingsworth, 2002; Imants & Van der Wal, 2020), the timing of measurement remains a critical issue. Additionally, many intervention studies included in the analysis evaluated programs that were more intensive than what is typically implemented in practice (Cramer et al., 2019; Donath et al., 2023), which may further contribute to inflated effect estimates. These factors may also have influenced the model calculations in this meta-analysis, potentially distorting the estimated strength of associations between variables. Although z-transformation could help mitigate the impact of large effect size variations across studies, it is currently not possible to apply z-transformation within the OMASEM framework (Jak & Cheung, 2020), which represents a limitation. Therefore, it is essential that the proposed models are further tested using high-quality primary data specifically designed to capture long-term outcomes and more realistic training settings. To mitigate these concerns, we repeated the model analyses using only data from studies with randomized or quasi-experimental designs. These analyses again supported the Willingness-to-Change model and also indicated acceptable fit for the Knowledge-Skills-Beliefs model. Overall, these limitations highlight the need for future primary research to employ robust study designs and comprehensive data collection, thereby validating and refining the tested models.

A fourth potential concern is slight study bias, although the bias appears to manifest in an unexpected direction in this case. In the present study, the higher power of the studies was related to the longer duration of the investigated teacher training, which might explain the observed bias. No differences were observed between published and unpublished studies. These factors underscore the need for further primary research, incorporating comprehensive data collection and robust study designs, to validate and refine the tested models.

Conclusion

Future research should integrate established theories on transfer processes when designing and evaluating teacher training programs (Ahadi et al., 2024), with the present study suggesting the Willingness-to-Change model as a promising framework for this purpose. Aligning research designs and professional development models more closely with teacher motivation and learning motivation can lead to deeper insights and more effective training outcomes (Kennedy, 2016). All models discussed in this study share that knowledge acquisition is necessary for the change process (Sims et al., 2023). However, on its own, it is insufficient to predict successful transfer. Knowledge may fail to drive lasting change if it cannot be meaningfully integrated with prior understanding or fully grasped to make classroom application feasible (Gregoire, 2003; Sims et al., 2023). Therefore, training effectiveness should be evaluated across multiple teacher-related dimensions (Ahadi et al., 2024). Future research should address aspects of teacher agency, especially teachers' willingness to change and their self-efficacy for IE, as these may critically mediate whether and how training content translates into practice.

As discussed in the limitations, future studies might address issues of study quality, including valid understandings of training impact, qualitative measures of variables relevant to assessing training effectiveness, and especially in long-term designs that capture effects at both the teacher and student levels (Nīmante et al., 2025). As highlighted by the low power of the primary studies in this MASEM study, large-scale, randomized, and quasi-experimental studies on teacher training remain scarce, particularly those that include student data (Kowalski et al., 2020). Since transfer processes are complex rather than linear, evaluating teacher training poses significant challenges (King et al., 2023). However, it is essential to include data collection at the student level to make meaningful statements on training success. If research continues to neglect conducting high-quality studies with large sample sizes at both the student and teacher levels, as well as comparisons with control groups, conclusions and review studies will remain limited by the lack of high-quality data.

Overall, this study demonstrated that complex processes underlie the transfer of knowledge acquisition to practice in teacher training, as well as improvements in students' behavior. Consequently, considerable future research is needed to further understand and support teachers and students in an educational system that is constantly facing changes implemented by political systems and technological advancements. Without a solid understanding of why and how teacher training can be effective, changes across the educational system may be implemented inefficiently at best and, at worst, be a mere paper-pushing exercise.

Declarations

References

Ahadi, A., Bower, M., Lai, J., Singh, A., & Garrett, M. (2024). Evaluation of teacher professional learning workshops on the use of technology: A systematic review. *Professional Development in Education, 50*(1), 221–237. https://doi.org/10.1080/19415257.2021.2011773

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Akin-Sabuncu, S., & Calik, B. (2023). A structural equation model of teachers' attitudes towards constructivist curriculum change. *International Journal of Curriculum and Instructional Studies, 13*(1), 123–157. https://doi.org/10.31704/ijocis.2023.006

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341–358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Azizatur, L., Irawati, S., & Lestari, N. P. (2021). The effect of training and development on teacher performance mediated by readiness to change during COVID-19 pandemic. *Jamanika (Jurnal Manajemen Bisnis Dan Kewirausahaan), 1*(3), 232–240.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*(1), 63–105.

Basma, B., & Savage, R. (2023). Teacher professional development and student reading in middle and high school: A systematic review and meta-analysis. *Journal of Teacher Education, 74*(3), 214–228. https://doi.org/10.1007/s10212-014-0241-6

Behrmann, L., & Souvignier, E. (2015). Effects of ft between teachers' instructional beliefs and didactical principles of reading programs. *European Journal of Psychology of Education, 30*(3), 295–312. https://doi.org/10.1007/s10212-014-0241-6

Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., et al. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal, 37*(3), 477–497. https://doi.org/10.1002/smj.2338

Bergren, B. A. (1997). *Teacher attitudes toward included special education students and co-teaching.* (ERIC Document Reproduction Services No. 408 754), U.S. Department of Education.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36*(4), 1065–1105. https://doi.org/10.1177/0149206309352880

Boeve-de Pauw, J., Olsson, D., Berglund, T., & Gericke, N. (2022). Teachers' ESD self-efficacy and practices: A longitudinal study on the impact of teacher professional development. *Environmental Education Research, 28*(6), 867–885. https://doi.org/10.1080/13504622.2022.2042206

Bonner, S. M., Diehl, K., & Trachtmann, R. (2019). Teacher belief and agency development in bringing change to scale. *Journal of Educational Change.* https://doi.org/10.1007/s10833-019-09360-4

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018). Rethinking models of professional learning as tools: A conceptual analysis to inform research and practice. *Professional Development in Education, 44*(1), 120–139. https://doi.org/10.1080/19415257.2017.1306789

Boylan, M., Adams, G., Perry, E., & Booth, J. (2023). Re-imagining transformative professional learning for critical teacher professionalism: a conceptual review. *Professional Development in Education, 49*(4), 651–669.

Boyle, C., Anderson, J., & Allen, K.-A. (2020). The importance of teacher attitudes to inclusive education. In C. Boyle, J. Anderson, A. Page, & S. Mavropoulou (Eds.), *Inclusive education: Global issues & controversies* (pp. 127–146). Brill. https://doi.org/10.1163/9789004431171_008

Cahapay, M. (2021). Kirkpatrick model: Its limitations as used in higher education evaluation. *International Journal of Assessment Tools in Education, 8*(1), 135–144. https://doi.org/10.21449/ijate.856143

Castillo, J. M., Wang, J. H., Daye, J. G., Shum, K. Z., & March, A. L. (2018). A longitudinal analysis of the relations among professional development, educators' beliefs and perceived skills, and response-to-intervention implementation. *Journal of Educational and Psychological Consultation, 28*(4), 413–444. https://doi.org/10.1080/10474412.2017.1394864

Chalmers, D., & Gardiner, D. (2015). An evaluation framework for identifying the effectiveness and impact of academic teacher development programmes. *Studies in Educational Evaluation, 46*, 81–91. https://doi.org/10.1016/j.stueduc.2015.02.002

Cheung, M. W. L. (2014). Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46*, 29–40. https://doi.org/10.3758/s13428-013-0361-y

Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review, 30*, 727–744. https://doi.org/10.1007/s10648-018-9437-7

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947–967.

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. *Journal for Research in Mathematics Education, 4*, 125–210.

Cohen, D. K., & Mehta, J. D. (2017). Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *American Educational Research Journal, 54*(4), 644–690. https://doi.org/10.3102/0002831217700078

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

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis, 42*(2), 208–231. https://doi.org/10.3102/0162373720906217

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

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.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective teacher professional development.* Palo Alto, CA: Learning Policy Institute.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession. *Washington, DC: National Staff Development Council, 12*(10), 1–33.

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

Desimone, L. M. (2023). Rethinking teacher PD: A focus on how to improve student learning. *Professional Development in Education, 4*(1), 1–3. https://doi.org/10.1080/19415257.2023.2162746

Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness, 13*(1), 29–66. https://doi.org/10.1080/19345747.2019.1670884

Doherty, I. (2011). Evaluating the impact of professional development on teaching practice: Research findings and future research directions. *US-China Education Review, 5*, 703–714.

Donath, J. L., Lüke, T., Graf, E., Tran, U. S., & Götz, T. (2023). Does professional development effectively support the implementation of inclusive education? *A Meta-Analysis. Educational Psychology Review, 35*(1), 1–30. https://doi.org/10.1007/s10648-023-09752-2

Dreer, B., Dietrich, J., & Kracke, B. (2017). From in-service teacher development to school improvement: Factors of learning transfer in teacher education. *Teacher Development, 21*(2), 208–224. https://doi.org/10.1080/13664530.2016.1224774

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

Eccles, J. S., & Wigfield, A. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research, 88*(3), 401–433. https://doi.org/10.3102/0034654317751918

Evans, D. K., & Popova, A. (2016). Cost-effectiveness analysis in development: Accounting for local costs and noisy impacts. *World Development, 77*, 262–276. https://doi.org/10.1016/j.worlddev.2015.08.020

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

Florian, L., & Camedda, D. (2020). Enhancing teacher education for inclusion. *European Journal of Teacher Education, 43*(1), 4–8. https://doi.org/10.1080/02619768.2020.1707579

Fogaça, J. L., Quartiroli, A., & Wagstaff, C. R. (2024). Professional development of sport psychology practitioners: From systematic review to a model of development. *Psychology of Sport and Exercise, 70*, Article 102550. https://doi.org/10.1016/j.psychsport.2023.102550

Gaastra, G. F., Groen, Y., Tucha, L., & Tucha, O. (2020). Unknown, unloved? Teachers' reported use and effectiveness of classroom management strategies for students with symptoms of ADHD. *Child & Youth Care Forum, 49*, 1–22. https://doi.org/10.1007/s10566-019-09515-7

Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review, 29*(3), 432–439. https://doi.org/10.1016/j.econedurev.2009.10.002

Gouëdard, P. Pont, B., Hyttinen, S. & Huang, P. (2020). Curriculum reform: A literature review to support effective implementation. *OECD Education Working Papers No. 239.* https://doi.org/10.1787/efe8a48c-en

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

Grohmann, A., Beller, J., & Kauffeld, S. (2014). Exploring the critical role of motivation to transfer in the training transfer process. *International Journal of Training and Development, 18*(2), 84–103. https://doi.org/10.1111/ijtd.12030

Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development, 15*(2), 103–120. https://doi.org/10.1111/j.1468-2419.2011.00373.x

Gülsün, I., Malinen, O. P., Yada, A., & Savolainen, H. (2023). Exploring the role of teachers' attitudes towards inclusive education, their self-efficacy, and collective efficacy in behaviour management in teacher behaviour. *Teaching and Teacher Education, 132*, 104–228. https://doi.org/10.1016/j.tate.2023.104228

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education, 21*(8), 967–983. https://doi.org/10.1016/j.tate.2005.06.007

Herawati, R., Tjahjono, H. K., Qamari, I. N., & Wahyuningsih, S. H. (2022). Teachers' willingness to change in adapting to online learning during the covid-19 pandemic. *Jurnal Cakrawala Pendidikan, 41*(2), 425–436. https://doi.org/10.21831/cp.v41i2.43233

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ignatavicius, D., & Chung, C. E. (2016). Professional development for nursing faculty: Assessing transfer of learning into practice. *Teaching and Learning in Nursing, 11*(4), 138–142. https://doi.org/10.1016/j.teln.2016.05.005

Imants, J., & Van der Wal, M. M. (2020). A model of teacher agency in professional development and school reform. *Journal of Curriculum Studies, 52*(1), 1–14. https://doi.org/10.1080/00220272.2019.1604809

Jak, S., & Cheung, M.W.-L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430–455. https://doi.org/10.1037/met0000245

Jak, S., Li, H., Kolbe, L., de Jonge, H., & Cheung, M. W. L. (2021). Meta-analytic structural equation modeling made easy: A tutorial and web application for one-stage MASEM. *Research Synthesis Methods, 12*(5), 590–606. https://doi.org/10.1002/jrsm.1498

Joyce, B. (1987). Staff development and student learning: A synthesis of research on models of teaching. *Educational Leadership, 45*(2), 11–23.

Kadluba, A., Strohmaier, A., Schons, C., & Obersteiner, A. (2024). How much C is in TPACK? A systematic review on the assessment of TPACK in mathematics. *Educational Studies in Mathematics.* https://doi.org/10.1007/s10649-024-10357-x

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. *Journal of In-service Education, 31*(2), 235–250. https://doi.org/10.1080/13674580500200277

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800

King, F., Poekert, P., & Pierre, T. (2023). A pragmatic meta-model to navigate complexity in teachers' professional learning. *Professional Development in Education, 49*(6), 958–977. https://doi.org/10.1080/19415257.2023.2248478

Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels.* Berrett-Koehler Publishers.

Knowles, M. (1984). *Andragogy in action.* Jossey-Bass.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

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

Kowalski, S. M., Taylor, J. A., Askinas, K. M., Wang, Q., Zhang, Q., Maddix, W. P., & Tipton, E. (2020). Examining factors contributing to variation in effect size estimates of teacher outcomes from studies of science teacher professional development. *Journal of Research on Educational Effectiveness.* https://doi.org/10.1080/19345747.2020.1726538

Krippendorff, K. (2011). *Computing Krippendorff's alpha reliability.* Retrieved from https://repository.upenn.edu/asc_papers/4. Accessed 14 Sept 2024.

Kuok, A. C., Teixeira, V., Forlin, C., Monteiro, E., & Correia, A. (2022). The effect of self-efficacy and role understanding on teachers' emotional exhaustion and work engagement in inclusive education in Macao (SAR). *International Journal of Disability, Development and Education, 69*(5), 1736–1754. https://doi.org/10.1080/1034912X.2020.1808949

Lazdina, S., & Daga-Krumina, E. (2023). Teachers' beliefs about teaching and learning: Why is it still a challenge? *Proceedings of ATEE Annual Conference, Latvia,* 272–291. https://doi.org/10.22364/atee.2022.17

Lieberman, A., & Miller, L. (1990). Teacher development in professional practice schools. *Teachers College Record, 92*(1), 105–122.

Lindner, K. T., Schwab, S., Emara, M., & Avramidis, E. (2023). Do teachers favor the inclusion of all students? A systematic review of primary schoolteachers' attitudes towards inclusive education. *European Journal of Special Needs Education, 38*(6), 766–787. https://doi.org/10.1080/08856257.2023.2172894

Lipowsky, F. (2010). Lernen im Beruf-Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. *Lehrerinnen und Lehrer lernen. Konzepte und Befunde Zur Lehrerfortbildung, 1*, 51–72.

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione, 7*(2), 27–51.

Loyalka, P., Popova, A., Li, G., & Shi, Z. (2019). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics, 11*(3), 128–154. https://doi.org/10.1257/app.20170226

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. Accessed 16 Sept 2024.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research, 85*(3), 353–394. https://doi.org/10.3102/0034654314553127

Meiers, M., & Ingvarson, L. (2005). *Investigating the links between teacher professional development and student learning outcomes.* Department of Education, Science and Training.

Moon, B. (2004). The open learning environment: A new paradigm for international developments in teacher education. In M. Ben-Peretz, S. Brown, & B. Moon (Eds.), *Routledge International Companion to Education* (pp. 756–770). Routledge.

Moore, D. E., Jr., Green, J. S., & Gallis, H. A. (2009). Achieving desired results and improved outcomes: Integrating planning and assessment throughout learning activities. *Journal of Continuing Education in the Health Professions, 29*(1), 1–15. https://doi.org/10.1002/chp.20001

Nīmante, D., Kokare, M., Baranova, S., & Surikova, S. (2025). Transferring results of professional development into practice: A scoping review. *Education Sciences, 15*(1), Article 95. https://doi.org/10.3390/educsci15010095

Opfer, D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*, 376–407. https://doi.org/10.3102/0034654311413609

Osman, D. J., & Warner, J. R. (2020). Measuring teacher motivation: The missing link between professional development and practice. *Teaching and Teacher Education, 92*, Article 103064. https://doi.org/10.1016/j.tate.2020.103064

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer, 37*(1), 107–136. https://doi.org/10.1093/wbro/lkab006

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research, 109*(5), 491–502. https://doi.org/10.1080/00220671.2014.992581

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 the American Medical Association, 298*(9), 1002–1009.

Reinhold, S., Gegenfurtner, A., & Lewalter, D. (2018). Social support and motivation to transfer as predictors of training transfer: Testing full and partial mediation using meta-analytic structural equation modelling. *International Journal of Training and Development, 22*(1), 1–14. https://doi.org/10.1111/ijtd.12115

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. John Wiley & Sons.

Rutherford, T., Long, J. J., & Farkas, G. (2017). Teacher value for professional development, self-efficacy, and student outcomes within a digital mathematics intervention. *Contemporary Educational Psychology, 51*, 22–36. https://doi.org/10.1016/j.cedpsych.2017.05.005

Salameh, J.-P., Bossuyt, P. M., McGrath, T. A., Thombs, B. D., Hyde, C. J., Macaskill, P., Deeks, J. J., Leeflang, 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

Sancar, R., Atal, D., & Deryakulu, D. (2021). A new framework for teachers' professional development. *Teaching and Teacher Education, 101*, 103–305. https://doi.org/10.1016/j.tate.2021.103305

Sasson, I., & Miedijensky, S. (2023). Transfer skills in teacher training programs: The question of assessment. *Professional Development in Education, 49*(2), 243–256. https://doi.org/10.1080/19415257.2020.1839783

Scott, K., Dawson, G., & Quach, J. (2024). How are we measuring domains that influence teacher readiness for change? A scoping review of existing instruments in non-tertiary settings. *Journal of Educational Change*. https://doi.org/10.1007/s10833-024-09521-0

Shaha, S. H., Glassett, K. F., & Copas, A. (2015). The impact of teacher observations with coordinated professional development on student performance: A 27-state program evaluation. *Journal of College Teaching and Learning, 12*(1), 55–64.

Sharma, U., & Mannan, H. (2015). Do attitudes predict behaviour: An (un) solved mystery?. In C. Forlin, S. Danford, & P. Jones (Eds.), *Foundations of inclusive education research* (pp. 115–131). Emerald Group Publishing Limited.

Shulman, L. S., & Shulman, J. H. (2004). How and what teachers learn: A shifting perspective. *Journal of Curriculum Studies, 36*(2), 257–271. https://doi.org/10.1080/0022027032000148298

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., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Van Herwegen, J., & Anders, J. (2023). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research*. https://doi.org/10.3102/00346543231217480

Somma, M. (2020). From segregation to inclusion: Special educators' experiences of change. *International Journal of Inclusive Education, 24*, 381–394. https://doi.org/10.1080/13603116.2018.1464070

Steele, L. M., Mulhearn, T. J., Medeiros, K. E., Watts, L. L., Connelly, S., & Mumford, M. D. (2016). How do we know what works? A review and critique of current practices in ethics training evaluation. *Accountability in Research, 23*(6), 319–350. https://doi.org/10.1080/08989621.2016.1186547

Strang, J. (2022). *Enhancing formative assessment through a professional development program: An examination of teacher change* (Doctoral dissertation, Piedmont University). https://www.proquest.com/openview/619f0a457d6a88baded614c508073873/1?pq-origsite=gscholar&cbl=18750&diss=y

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education, 67*, 291–301. https://doi.org/10.1016/j.tate.2017.06.020

Sztajn, P., Campbell, M. P., & Yoon, K. S. (2011). Conceptualizing professional development in mathematics: Elements of a model. *PNA, 5*(3), 83–92.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2008). *Teacher professional learning and development* (Vol. 18). International Adacemy of Education.

Voelkel, R. H., Jr., & Chrispeels, J. H. (2017). Within-school differences in professional learning community effectiveness: Implications for leadership. *Journal of School Leadership, 27*(3), 424–451. https://doi.org/10.1177/105268461702700305

Wade, R. K. (1985). What makes a difference in in-service teacher education? A meta-analysis of research. *Educational Leadership, 42*, 48–54.

Wu, S. Y. (2021). How teachers conduct online teaching during the COVID-19 pandemic: A case study of Taiwan. *Frontiers in Education, 6*, Article 675434. https://doi.org/10.3389/feduc.2021.675434

Yada, A., Tolvanen, A., & Savolainen, H. (2018). Teachers' attitudes and self-efficacy on implementing inclusive education in Japan and Finland: A comparative study using multi-group structural equation modelling. *Teaching and Teacher Education, 75*, 343–355. https://doi.org/10.1016/j.tate.2018.07.011

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., Gomez, C. J., Moreno, L., Rolla, A., D'Sa, N., & Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51*(3), 309. https://doi.org/10.1037/a0038785

Zambak, V. S., Alston, D. M., Marshall, J., & Tyminski, A. M. (2017). Convincing science teachers for inquiry-based instruction: Guskey's staff development model revisited. *Science Educator, 25*(2), 108–116.