



Facilitating Diagnostic Competences in Higher Education—a Meta-Analysis in Medical and Teacher Education

Olga Chernikova¹  · Nicole Heitzmann^{1,2} · Maximilian Christian Fink² · Venance Timothy¹ · Tina Seidel³ · Frank Fischer¹ · DFG Research group COSIMA

Published online: 3 July 2019
© The Author(s) 2019

Abstract

Facilitating diagnostic competences is an important objective of higher education for many professions. This meta-analysis of 35 empirical studies builds on a conceptual framework and investigates the role of problem-solving, scaffolding, and context to foster diagnostic competences in learners with lower and higher professional knowledge bases. A moderator analysis investigates which type of scaffolding is effective for different levels of learners' knowledge bases, as well as the role of the diagnostic context. Instructional support has a moderate positive effect ($g = .39$; CI [.22; .56]; $p = .001$). Diagnostic competences are facilitated effectively through problem-solving independent of the learners' knowledge base. Scaffolding types providing high levels of guidance are more effective for less advanced learners, whereas scaffolding types relying on high levels of self-regulation are more effective for advanced learners.

Keywords Diagnostic competences · Scaffolding · Teacher education · Medical education · Meta-analysis

Olga Chernikova and Nicole Heitzmann shared first authorship.

✉ Olga Chernikova
o.chernikova@psy.lmu.de

¹ Ludwig-Maximilian University of Munich (LMU), Leopoldstr. 13, 80802 Munich, Germany

² Institute of Medical Education, University Hospital, Ludwig-Maximilian University of Munich (LMU), Munich, Germany

³ Friedl Schöller Endowed Chair of Educational Psychology, Technical University of Munich (TUM), Munich, Germany

Making efficient decisions in professional fields is impossible without being able to identify, understand, and even predict situations and events relevant to the profession. Therefore, diagnosis is an essential part of professional competences in different domains. It involves problem identification, analysis of context, and application of obtained knowledge and experience to make practical decisions. The two fields of medical and teacher education specifically focus on the processes of collecting and integrating case-specific information to reduce uncertainty and make practical decisions. Teacher education deals with teachers' assessments of students' knowledge and learning processes, and medical education investigates primarily clinical reasoning to diagnose patients' diseases accurately. Despite these different professional contexts and relevant situations, the diagnostic processes and underlying competences required to come to medical or educational decisions are similar. This similarity has resulted in a call to explore a closer link between the two research traditions (e.g., Gartmeier et al. 2015; Stürmer et al. 2016).

Although quite strong empirical evidence supports learning through problem-solving in postsecondary education in general (Belland et al. 2017; Dochy et al. 2003), studies of the use of problem-solving for advancing diagnostic competences in medical and teacher education remain open and need more synthesized systematic evidence. Some empirical studies in medical and teacher education indicate positive effects of additional scaffolding (e.g., structured reflection) in diagnostics-related instruction (Ibiapina et al. 2014; Klug et al. 2016; Mamede et al. 2014). However, other studies report no added value of scaffolding or even negative effects (Heitzmann et al. 2018a; Heitzmann et al. 2015; Stark et al. 2011). This variability in effects leads to further open questions. On the one hand, there are questions related to the optimal use of scaffolding to facilitate diagnostic competences. On the other hand, there are questions about the role of other factors, such as the nature of a diagnostic situation or prior professional knowledge of learners, which can also influence the outcomes.

This meta-analysis aims at providing answers to these questions and enhancing the scientific understanding of various factors and conditions (context, instructional, or personal) that facilitate diagnostic competences in the fields of medical and teacher education. Moreover, this meta-analysis contributes to identifying most effective scaffolding procedures to support learning through solving problems, depending on the levels of prior professional knowledge the learners have already acquired. This contribution provides insights for educators regarding the design and use of learning environments to enhance the advancement of professional diagnostic competences.

Diagnostic Competences in Medical and Teacher Education

Medical diagnosis aims at finding the cause of a disease and the appropriate courses of action for either further diagnosis or treatment (Charlin et al. 2000). Diagnostic processes in medical education focus on examining patients' body functioning, identifying pathological processes and possible risk factors, and preparing decisions about the most appropriate treatment. Diagnosing in teacher education aims at optimizing the use of instructional methods to close the gap between the present and desired states of the learners' competences (Helmke et al. 2012). Diagnostic processes in teacher education focus on examining students' characteristics relevant for learning and performance (e.g., motivation, intelligence), defining students' academic achievement and performance, and analyzing classroom situations, the impact of instruction, and contextual factors (Glogger-Frey et al. 2018). More generally, diagnosing first

focuses on comparing the current state of learners' knowledge and skills to predefined learning objectives, and subsequently aims at identifying misconceptions, difficulties, or particular needs of learners to choose the most appropriate instructional support to meet both learners' needs and learning objectives. While acknowledging the differences in diagnostic processes between the two fields, also obvious are shared commonalities—namely, that diagnosing a patient's health status or a learner's understanding is a goal-oriented process of collecting and integrating case-specific information to reduce uncertainty in order to make medical or educational decisions (Heitzmann et al. 2015).

Recent conceptualizations further suggest that accurate and effective diagnosing requires advanced diagnostic competences. These entail the coordinated application of different types of knowledge (e.g., Shulman 1987; Stark et al. 2011) relevant for professional diagnostic problems as well as following particular steps that lead to diagnostic decisions. Research on diagnostic competences (Heitzmann et al. 2018b) emphasizes the importance of learners' characteristics (i.e., prior professional knowledge base) and the diagnostic processes taking place during learning and assessment, and suggests indicators for the quality of processes and outcomes, as discussed in the following sections.

Medical education and teacher education entail different professional contexts and relevant situations. The main task of a teacher is to support the learning of an individual in a class. A medical doctors' main task is to support individuals to achieve and sustain good health. Another difference is the conceptual professional knowledge base of doctors and teachers. Nonetheless, commonalities in the professional practice of teachers and physicians also can be found: In both professions, decision-making is based on characteristics of other people's education or health, respectively. Interventions should be based on accurate diagnosis of a current state, often to infer causes of problems or future potentials. Therefore, different types of professional knowledge are coordinated in different diagnostic activities, such as generating hypotheses and evidence, drawing conclusions, and communicating the results. In addition, in medical as well as in teacher education, a main focus is the integration of case-specific information to reduce uncertainty and thus make practical decisions. Higher education programs aim to provide their students with hands-on participation and reflection on practical experience with the goal of advancing students' competences (e.g., Grossman et al. 2009).

Conceivably, the effects of instructional interventions that aim at facilitating diagnostic competences may be similar in cognitively similar situations that demand equal diagnostic activities across domains compared to cognitively dissimilar situations within the same domain. An instructional intervention that guides learners to generate hypotheses early and gives instruction on how to prioritize evidence might be beneficial for both prospective teachers and prospective medical practitioners.

Even though we are aware of the differences between teacher and medical education, we also point to the commonalities when proposing to link research in those domains to improve our understanding of the domain-general and domain-specific aspects of instructional interventions to facilitate diagnostic competences.

Mainly, the commonalities, together with different empirical research traditions, have resulted in a call to explore the possibilities of a closer link between the two domains and the respective research traditions (e.g., Gartmeier et al. 2015; Stürmer et al. 2016; Trempler et al. 2015). A question, then, might ask why we study only these two domains and not others with professional practices in which diagnostic processes are probably relevant as well, such as car mechanics or law. Although we believe this would be a promising new route of research, not enough of a joint conceptual basis to expand the comparison domains currently exists.

Even more important in the context of this meta-analysis, little empirical research and even less experimental work on scaffolding for the development of diagnostic competences has been pursued in these other fields.

Professional Knowledge as a Prerequisite for Advancing the Diagnostic Competences

Professional knowledge (together with cognitive skills and motivational factors) is one of the essential facets of competence and therefore one of the most important learning outcomes of professional training (Blömeke et al. 2015). The boundary approach to competence (Stoof et al. 2002) emphasizes that competence is more than the sum of domain-specific knowledge, skills, and motivational factors; these building blocks are strongly interconnected, but have different weights at different stages of learning and competence development. Nevertheless, the common way to assess level of competence is through assessing the abovementioned building blocks. Professional knowledge is the most commonly addressed component. It is measured frequently and objectively during assessment phases (both as a conceptual knowledge in written and oral tests but also as practical knowledge and knowledge application, measured at the level of skills). At the same time, professional knowledge is a fundamental prerequisite for further development of skills and competences. Professional knowledge defines the capacity to learn from different learning materials and instructional support (i.e., learners with low levels of prior knowledge might require more instructional support and guidance than advanced learners), which therefore might influence the choice of instructional approaches.

In the domain of medical education, the professional knowledge required for diagnosing has been differentiated into (a) biomedical knowledge, operationalized as knowledge about normal functioning and pathological processes in causing the disease (Kaufman et al. 2008), and (b) clinical knowledge, including knowledge about symptoms, symptom patterns, factors indicating high likelihood of particular diseases, and knowledge about appropriate treatment (Van de Wiel et al. 2000). Other researchers, such as Stark et al. (2011), suggest a more general differentiation of knowledge into conceptual knowledge (interrelation of terms) and practical knowledge (knowing how to apply conceptual knowledge to cases). They also further divide practical knowledge into two components: strategic (knowledge of problem-solving steps) and conditional (knowledge of conditions for successful application of problem-solving steps).

In the domain of teacher education, Shulman (1987) has proposed distinguishing among different teacher knowledge facets: content, pedagogical content, and pedagogical. *Content knowledge* is operationalized as knowledge of subject matter (e.g., division rules in arithmetic, photosynthesis processes in biology). *Pedagogical content knowledge* includes both knowledge of content (“what”) and knowledge of pedagogical principles to deliver this content (“how”), as well as typical misconceptions and explanations. *Pedagogical knowledge* is generic across particular domains and includes knowledge about memory, learning, and motivation function; classroom management; and general teaching strategies. In Shulman’s conception, teacher knowledge comprises aspects of acquiring conceptual knowledge regarding these three facets, as well as practical aspects with regard to acting in situations relevant to teaching. The same perspective is taken by Hiebert et al. (2002), for example, when proposing a professional knowledge base for teacher education that incorporates integrated conceptual and practical knowledge.

Although clear differences exist between the professional knowledge bases required in the fields of medical and teacher education (e.g., knowledge of symptoms of a disease versus knowledge about common misconceptions in math, or going through patient examination checklists versus formulating questions for a test), the distinction into conceptual and practical aspects seems to be a common denominator that can therefore be used as cross-domain concept of professional knowledge. Important open questions involve how the different types of knowledge are integrated and applied in decision-making and problem-solving contexts and whether instructional support can facilitate the acquisition of different knowledge types with similar effectiveness.

In summarizing the research in this field, it seems rather important to consider the possible effects of prior professional knowledge as a prerequisite for training and advancing diagnostic competences through means of instruction. In addition, studying differential effects regarding the advancement of conceptual *and* practical knowledge as learning outcomes is necessary. Both aspects are addressed in this meta-analysis. Conceptual and practical professional knowledge in this meta-analysis refers to the content of measured domain-specific knowledge, what the knowledge is about. The current research is not specific enough to infer the type of knowledge representations—declarative or procedural—suggested by ACT-R (Anderson et al. 2004).

Learning Environment and Learning Processes

Diagnostic processes have been conceptualized as a set of epistemic activities (Fischer et al. 2014), including (a) identifying a problem, (b) questioning, (c) generating hypotheses, (d) constructing artifacts, (e) generalizing evidence, (f) evaluating evidence, (g) drawing conclusions, and (h) communicating process and results. Facilitating these activities during learning phases seems essential for the advancement of competences because it provides multiple opportunities for learners to engage in various diagnostic practices. According to theories of expertise and skill development (e.g., Van Lehn 1996), practice opportunities, combined with sufficient professional knowledge, can facilitate diagnostic competences in higher education.

The training of diagnostic competences that include the application of epistemic-diagnostic activities has been discussed by a number of researchers in the light of complex problem-solving. A strong body of evidence shows that problem-centered approaches, such as problem-based learning, case-based learning, or learning through problem-solving, are effective instructional approaches for facilitating such skill-related outcomes as diagnostic competences in postsecondary education (Belland et al. 2017; Dochy et al. 2003).

However, exposure to complex and ill-structured problems, especially in the early stages of expertise development, is considered problematic from the perspective of cognitive load theory (Renkl and Atkinson 2003; Sweller 2005). According to this theory, learning by complex problem-solving should be effective for learners with knowledge sufficiently organized to enable self-regulated problem-solving, whereas learners missing these prerequisites might be overburdened. Contrary to this general claim from cognitive load theory about problem-centered approaches, Hmelo-Silver et al. (2007) regard problem-centered approaches as suitable for learners with little prior knowledge if high levels of *scaffolding* accompany the challenging tasks. Quintana et al. (2004) suggest that scaffolding enables a learner to achieve goals (i.e., solve problems) through modifying the task and reducing possible pathways from which to choose, and through prompts and hints to help the learner coordinate the steps in

problem-solving or interaction. In meta-analyses, scaffolding has shown positive effects on various learning outcomes, including complex competences (Belland et al. 2017; Devolder et al. 2012; Gegenfurtner et al. 2014).

Instructional Support for Advancing Diagnostic Competences

Guided Problem-Centered Instruction

In problem-centered instructional approaches, learners solve authentic cases with varying levels and types of instructional support (Belland et al. 2017). Problem-centered instructional approaches include, by definition, “problem-based learning, modeling/visualization, case-based learning, design-based learning, project-based learning, inquiry-based learning, [...] and problem solving” (Belland et al. 2017, p. 311). These instructional approaches frequently have been used in the past to facilitate diagnostic competences in medicine (Barrows 1996) and teacher education (Seidel et al. 2013). Theoretical arguments to advance diagnostic competences exist for the effectiveness of problem-solving (Anderson 1983; Jonassen 1997), case-based learning (Kolodner 1992), and problem-based learning (Barrows 1996). Previous meta-analytic results and reviews (Albanese and Mitchell 1993; Belland et al. 2017; Dochy et al. 2003; Thistlethwaite et al. 2012) support the effectiveness of these instructional approaches.

Scaffolding

The most prominent definition of scaffolding (Wood et al. 1976) considers it to be the process of supporting learners by regulating or limiting intricate factors of the task. This objective is accomplished by six *scaffolding functions*: (a) sparking situational interest, (b) reducing the complexity and difficulty of tasks, (c) keeping learners focused on their goal, (d) highlighting crucial features of a task, (e) motivating disappointed learners, and (f) providing solutions and models of a task. The concept of scaffolding builds on Vygotsky’s (1978) notion of the Zone of Proximal Development, which includes challenging tasks a learner can perform successfully with external guidance but would not yet be able to perform independently. According to recent literature reviews (Belland 2014; Reiser and Tabak 2014), the key components of scaffolding are formative assessment and adapting the level of support to the performance or prerequisites of the learner. Scaffolding can focus on cognitive, meta-cognitive, motivational, and strategic outcome measures (Hannafin et al. 1999).

In agreement with Belland (2014), we apply a comprehensive scaffolding definition that includes types of scaffolding with or without adapting support during the learning process (i.e., fading or adding support). Recent research on scaffolding in the context of learning through problem-solving suggests several techniques to structure and guide the facilitation of competences: (a) providing *examples*, which are partial or whole problem solutions or target behaviors (e.g., Renkl 2014); (b) providing *prompts*, or hints about how to handle materials or how to proceed with solving the problem (e.g., Quintana et al. 2004); (c) assigning *roles* to actively involve learners in learning tasks (e.g., Strijbos and Weinberger 2010); and (d) inducing *reflection phases*, which allow learners to think about goals of the procedure, analyze their own performance, and/or plan further steps (e.g., Mamede and Schmidt 2017). All possible scaffolding forms fall somewhere in a continuum from one in which only a specific

element of a scaffold is realized all the way to a full realization in which all elements are realized.

Providing Examples

In example-based learning, learners retrace the steps of a solution (worked example) or observe a model displaying the process of problem-solving (modeling example) before they solve problems independently (Renkl 2014). Example-based learning with worked and modeling examples has already been shown to be effective for the advancement of a variety of complex cognitive skills, such as scientific reasoning (Kirschner et al. 2006; Fischer et al. 2014) and scientific writing (Zimmerman and Kitsantas 2002), which possess some similarities to diagnostic competences. The *worked example effect* and the underlying cognitive load theory (Sweller 1994) suggest that problem-solving at early stages of knowledge or skill acquisition without scaffolding can lead to an excessive amount of information and an inhibition of schema acquisition (Renkl 2014). Worked examples are typically highly effective for beginners but reduced and even negative effects have been reported for intermediates (Van Gog and Rummel 2010). These effects, however, might be different for *modeling examples*, which are effective when learners possess sufficient prior knowledge to comprehend and evaluate the complex skills they observe (Van Gog and Rummel 2010). Our analysis would allow estimating the effects of providing examples to facilitate diagnostic competences and therefore extend the findings from complex cognitive skills to competences advancement.

Providing Prompts

Prompts refer to information or guidance offered to learners during the learning process in order to raise effectiveness (Berthold et al. 2007). Various types of prompts have differing objectives. *Self-explanation prompts* put an emphasis on the verbalization of reasoning and elaboration processes while solving a task (Heitzmann et al. 2015; Quintana et al. 2004). *Meta-cognitive prompts* raise awareness of meta-cognitive processes that control self-regulated learning (Quintana et al. 2004). *Collaboration scripts* assist the regulation of social interaction in interactive learning settings (Fischer et al. 2013; Vogel et al. 2017). The open questions are whether providing prompts significantly contributes to the advancement of diagnostic competences in teacher and medical education and whether the effects differ for students with lower and higher levels of prior professional knowledge.

Assigning Roles

Role taking can be considered as a type of scaffolding for which the full complexity of a situation is reduced by assigning a specific role with limited tasks or perspective on the full task. In teacher education, teacher and student are typical roles; in medical encounters, doctor and patient are typical. Additionally, learners can be assigned the role of observer. A large body of empirical research suggests that complex skills can be acquired effectively in the agent (i.e., teacher or doctor) role (Cook 2014). Results on acquiring diagnostic competences in the role of the observer are still lacking, but Stegmann et al. (2012) showed that communication skills can be acquired as effectively as in the agent role. Even though systematic research on the acquisition of diagnostic competences in the roles of patient and student is still lacking, it seems likely that learners may gain specific diagnostic competences and knowledge through

displaying clinical symptoms or student's mistakes and behaviors. Apart from the described results and mechanisms, findings on differences between beginners and intermediates are currently lacking.

Inducing Reflection Phases

The positive effects of reflection on learning were first proposed by Dewey (1933). In modern days, a comprehensive definition of reflection is best stated by Nguyen et al. (2014): "Reflection is the process of engaging the self in attentive, critical, exploratory and iterative interactions with one's thoughts and actions, and their underlying conceptual frame, with a view to changing them and with a view on the change itself" (p. 1182). Reflection can be induced through guided reflection phases and can take place before, during, or after an event. Reflection can occur in a social context (Nguyen et al. 2014) or individually (e.g., by writing reflective journals (O'Connell and Dymont 2011)). Different types of reflection have been reported to efficiently foster the acquisition of diagnostic competences in medicine (Sandars 2009) and in teacher education (Beauchamp 2015).

Reflection could facilitate diagnostic competences for three major reasons. First, reflection phases add an extra pause for the learner. Beginners might use this pause to better retrieve and apply conceptual knowledge with less time pressure (Renkl et al. 1996). Advanced learners might benefit significantly by having time not only to activate and better integrate the conceptual knowledge and previous experience but also to evaluate the selected strategy and think about alternatives. Second, learners may self-generate feedback internally which advances their learning during reflection (Butler and Winne 1995; Nicol and Macfarlane-Dick 2006). Third, reflection may support planning subsequent steps of the diagnostic process. Current meta-analysis would allow estimating the effects of the reflection in fostering diagnostic competences of learners with high and low levels of prior professional knowledge.

Scaffolding and Self-Regulation

A convincing framework to integrate the different types of scaffolding does not yet exist. Heuristically, we suggest building on the very idea of scaffolding as a temporary shift of control over the learning process from a learner to a teacher or more advanced peer (e.g., Tabak and Kyza 2018). We further suggest locating the scaffolding types on different positions on a scale of self-regulation of problem-solving, with examples located from a rather low level followed by role assignments and prompts with increasing levels of self-regulation potential. Reflection phases followed by unscaffolded problem-solving would be located at the high end of the self-regulation scale.

Although each approach adopts an idea of the transition from other-regulation to self-regulation at some stage (e.g., fading of steps in worked examples, internalizing collaboration scripts), the suggested classification allows estimation of the amount of content support and guidance initially provided through that type of scaffolding by conceptualizing scaffolding types into a continuous dimension with increasing degrees of freedom for the learner. Similar ideas, such as classifying scaffolding measures based on the amount and kind of guidance, were introduced by Brush and Saye (2002). These authors suggested a dichotomous categorization into (a) "soft" scaffolding, focused on fostering meta-cognitive skills and self-regulation, and (b) "hard" scaffolding, focused on content: conceptual or procedural knowledge required to solve the task, providing learners with full or partial solutions to foster learning.

Brush and Saye (2002) claim that “hard” scaffolding is beneficial for initial stages of learning, whereas “soft” scaffolding is more beneficial when initial knowledge is already acquired.

Context Factors in Facilitating Diagnostic Competences

The characteristics of real-life situations in which the diagnostic processes take place are important considerations in any facilitation. For the *domain* of medical education, the factors include identifying the cause of the disease to further plan treatment steps; for the *domain* of teacher education, they include assessing students’ knowledge level and identifying misconceptions to adjust teaching strategy accordingly or to suggest additional support.

With regard to the *nature of the diagnostic situation*, we adopt the classification by Heitzmann et al. (2018b), who distinguish two dimensions. The first dimension is *information base* (e.g., where the information for diagnosis comes from), spanning the spectrum from document-based to interaction-based. Document-based diagnosis relies on information available in written or otherwise recorded form (laboratory findings, x-ray images, students’ academic achievement scores, students’ homework). There is little or no time pressure for the analysis of this information, which can be accessed several times if needed with reflection always possible. Interaction-based diagnosis relies on the information received through communication with patients, students, or their families (e.g., anamnestic interview, oral exam, teacher-guided in-class discussions). Information from interaction usually needs to be processed in “real time,” which involves more time pressure and fewer opportunities for reflection.

The second dimension to describe diagnostic situations according to Heitzmann et al. (2018b) is on a continuum from individual diagnoses to the necessity for *collaboration* and communication with other professionals during the diagnostic processes. Empirical studies provide evidence that collaboration and social aspects of working on the case can be problematic even for experts, requiring additional knowledge and skills (e.g., Kiesewetter et al. 2013).

This meta-analysis uses context factors to address (a) the role of the diagnostic situation in organizing learning processes to facilitate diagnostic competences and (b) the generalizability of findings across domains.

Interaction Between Professional Knowledge Base and Instructional Support

The professional knowledge base of the learner informs the requirements for the organization of learning processes, choice of learning and teaching strategies, and the amount and type of guidance (Renkl and Atkinson 2003; Sweller 2005; Van Lehn 1996). It is therefore essential to explore the effectiveness of different instructional support measures in relation to the learner’s prior knowledge base. This determination would contribute to practical considerations in the development of educational programs, which addresses the lack of empirical evidence concerning the role of different types of scaffolding for learners with lower and higher levels of prior professional knowledge in facilitating diagnostic competences.

The theoretical framework and the categorization of scaffolding procedures into a continuum based on the degree of self-regulation of problem-solving assumes that

learners who already have sufficient conceptual knowledge and experience in its application would benefit more from guidance that allows higher degrees of freedom and the opportunity for more self-regulation, such as by introducing reflection phases built into the problem-solving process. Learners who do not have sufficient prior conceptual or procedural knowledge in the field are expected to benefit more from higher levels of the types of scaffolding that provide them with conceptual knowledge and heuristics for decision-making, such as instructional support that provides examples (e.g., professional solutions, worked-out examples, or behavioral models) or assigns roles. This meta-analysis aims at generating evidence on whether the professional knowledge base moderates the effects of scaffolding types on advancement of diagnostic competences.

Research Questions

RQ1: To what extent can instructional support facilitate diagnostic competences in higher education?

We assume that, in line with research on learning (Hmelo-Silver et al. 2007; Kirschner et al. 2006; Renkl and Atkinson 2003; Van Lehn 1996) and previous meta-analyses on the effects of instructional support on acquisition of complex cognitive skills and competences (Belland et al. 2017; Devolder et al. 2012; Dochy et al. 2003; Gegenfurtner et al. 2014), instructional support would have a positive effect on development of diagnostic competences in medical as well as in teacher education.

RQ2: What is the role of professional knowledge in the acquisition of diagnostic competences?

We assume that prior professional knowledge would be a significant moderator of the effects of instructional support to facilitate diagnostic competences, as professional knowledge is an essential part of professional competence (Blömeke et al. 2015; Stoof et al. 2002). In general, we assume that instructional support would have higher effects on diagnostic competences of learners with low levels of prior knowledge.

RQ3: How do problem-solving and different types of scaffolding facilitate diagnostic competences?

We assume that introducing elements of problem-solving, as instructional approach as well as different types of scaffolding, would all have positive yet different effects on the advancement of diagnostic competences.

RQ4: To what extent do effects of learning through problem-solving and scaffolding depend on professional knowledge?

We assume that the level of prior professional knowledge will moderate the effects of the scaffolding. We further assume that learners with high prior knowledge would benefit more from the types of scaffolding requiring higher levels of self-regulation required, whereas learners with low levels of prior knowledge would benefit more from the types of scaffolding with lower levels of self-regulation and more guidance (Sweller 2005; Van Lehn 1996).

RQ5: What are the roles of contextual factors (i.e., domain, need to collaborate, source of information to use for diagnosing) in facilitating diagnostic competences?

We assume that contextual factors would affect the diagnostic competence advancement and therefore be significant moderators. The analysis should also provide evidence on the generalizability of the findings across the domains of medical and teacher education.

Method

Inclusion and Exclusion Criteria

The inclusion criteria were based on outcome measures reported, research design applied, and statistical information provided in the studies. We discuss these criteria below in more detail.

Diagnostic Competences

The studies eligible for inclusion had to focus on facilitation of diagnostic competences, defined as dispositions enabling goal-oriented gathering and integration of information in order to make medical or educational decisions. In particular, studies in teacher education must be related to measurement of professional vision or formative assessment. The outcome measures had to address either diagnostic quality and/or one or more of several epistemic-diagnostic activities (Fischer et al. 2014): identifying the problem, questioning, generating hypotheses, constructing artifacts, generalizing and evaluating evidence, drawing conclusions, and communicating processes and results. Any studies that did not include any epistemic-diagnostic activities or did not report measures of diagnostic activities (or professional vision or formative assessment), as well as studies that focused on acquisition of motor skills in medical education, were excluded from the analysis. This meta-analysis focuses only on objective measures of learning (written or oral knowledge tests, assessment of performance based on expert rating, or any quantitative measures, including but not limited to frequency of behavior or number of procedures performed correctly). Studies that reported only learners' attitudes, beliefs, or self-assessment of learning or competence were excluded from the analysis.

Research Design

The aim of this meta-analysis was to make causal inferences regarding the effect of instructional support on diagnostic competences, so the studies eligible for the analysis had to have an experimental design with at least one treatment and one control condition. The treatment condition had to include instructional support measures directed at facilitating diagnostic competences not included in the control condition. The studies that did not report any intervention (i.e., studies on tool or measurement validation), studies that reported comparison of multiple experimental designs (e.g., instruction with few prompts versus many prompts, or using best practices examples versus erroneous examples), and studies that did not provide any control condition, such as waiting condition or historical control, were excluded from the analysis.

Study Site, Language, and Publication Type

Eligible studies were not limited to any specific study site. To ensure that the concepts and definitions of the core elements coded for the meta-analysis were comparable and relevant, only studies published in English were included in the analysis. However, the origin of studies and language of conduction were not restricted. Different sources, both published and unpublished, were considered to ensure the validity and generalizability of the results. There were no limitations regarding publication year.

Effect Sizes

Eligible studies were required to report sufficient data (e.g., sample sizes, descriptive statistics) to compute effect sizes and identify direction of scoring. If a study reported the information about pretest effect size, it was used to adjust for pretest differences between treatment and control conditions.

Search Strategies

To perform the meta-analysis, the following databases were screened for eligible empirical studies: PsycINFO, PsycINDEX, PsycARTICLES, ERIC, and MEDLINE. The search terms used were (professional vision OR formative assessment* OR diagnost* competenc* OR diagnost* skill* OR diagnost* reason* OR clinical reason*) AND (train* OR teach*). Additionally, the first authors of eligible studies were contacted to obtain information about other published or unpublished manuscripts, and references were checked for other studies. The search results were obtained on February 24, 2018. The search resulted in 7510 documents (after deletion of duplicates).

Coding Procedures

The first phase involved screening. Screening for eligibility was conducted based on the inclusion/exclusion criteria mentioned above. A study was excluded from the analysis only if it provided enough information and met one or more exclusion criteria. For example, if the study title provided enough information to exclude the study, the study was excluded already at this stage. If information was insufficient for exclusion, the study was included in further screening (abstract or full text). The work was shared between the fourth and one of the two first authors. The fourth author individually examined titles, abstracts, and full texts of the studies to identify those with the eligible search terms “diagnostic competence,” “diagnostic skill,” “diagnostic reasoning,” or “clinical reasoning.” Then, the first author of the study screened the abstracts and the full texts for the search terms “professional vision” and “formative assessment,” and marked studies that needed further examination. No interrater agreement was determined for the screening stage. However, in regular meetings, the authors of this meta-analysis discussed studies with insufficient information or complex study designs with respect to eligibility until complete agreement on inclusion or exclusion of a study was achieved.

The second phase (coding) used a previously piloted coding scheme, which was refined until a sufficient interrater reliability was achieved. For an overview of the coding manual, please see Appendix Tables 5, 6, 7, 8, and 9. Features of primary studies (study design, use of instructional support measures for treatment and control groups, professional knowledge base, and context) were independently double coded by one of the authors and a trained research assistant. The training procedure involved both individuals coding one of the studies together, and then each coder coded another study of the sample independently, followed by a discussion of the differences in the coding. The satisfactory interrater agreement of the subsequent ratings (above .75) pointed to the success of the training procedure. Remaining disagreements on the ratings were resolved in regular meetings of the coders until 100% agreement on all codes was achieved. The

data extracted from eligible primary studies included study characteristics, independent and dependent variables, and statistical values needed for calculating effect sizes.

Study Characteristics

Study characteristics extracted for each study included information about authors and publication year as well as information about sample size. Additionally, studies were coded for study design (i.e., random distribution of treatment and control condition). An overview of study characteristics and moderators is presented in Table 1. Some more descriptive statistics about primary studies including participants, measurements used, and a summary of results are presented in Table 2.

Coding for the Moderators

The coding scheme, including the professional knowledge base, instructional support and scaffolding, and diagnostic context, was based on a recent conceptual framework (Heitzmann et al. 2018b). The *domain* was coded as either medical or teacher education.

The *professional knowledge base* was coded as “low” if participants of the study had little or no exposure to similar context (i.e., no or low prior conceptual/procedural knowledge regarding the assignments during the learning phase) or were in the initial phase of their training (also indicating a rather low prior professional knowledge base on conceptual and procedural levels). It was coded as “high” if learners already had experience, were exposed to a similar context, or were in the final phase of their training (indicating that learners already had a high professional knowledge base on conceptual and procedural levels).

Instructional support was coded as “yes” (included) or “no” (not included) for the following categories:

- *Problem-solving* (using a problem-centered approach) was coded as “included” if learners received cases/problems and made diagnostic decisions themselves, or “not included.”
- *Examples* were coded as included if learners observed modeled behavior, example solutions, or worked examples at some time during the training, or not included.
- *Prompts* were coded as included if learners received hints on how to handle the learning material to support them in solving their diagnostic problem. Additionally, the type of prompts (time of exposure) were coded as “during” if prompts were provided during completion of the learning task before the diagnosis/decision, as “after” if support was given after the diagnosis/decision had been made, as “longterm” if the support was provided in various steps of a diagnostic process that takes place over a long period of time (e.g., reflective diaries, longitudinal studies), and as “mixed” if more than one type of prompt was present. This rather rough type of clustering has been chosen because the number of studies was insufficient for a more fine-grained analysis.
- *Roles* were coded as “included” if learners acted as physicians or teachers (“agent”), observers, patients, or students (“other”) at least part of the time, or as “not included.”
- *Reflection phases* were coded as included if learners were encouraged to think about the goals of the procedure, analyze their own performances, and plan further steps, or “not included.”

Table 1 Study characteristics

Authors, year	Instructional support							Diagnostic situation		
	Domain	Sample size (N)	Knowledge base	Problem-solving	Examples	Prompts	Roles	Reflection phases	Info base	Processing mode
1 Baghdady et al. 2014	Medical	112	Low	No	No	No	No	No	DOC	Individual
2 Bahreini et al. 2013	Medical	73	High	Yes	No	Longterm	Agent_full	Yes	INT	Individual
3 Besser et al. 2015	Teacher	67	High	Yes	Yes	Longterm	Agent_part	Yes	INT	Collaboration
4 Chamberland et al. 2011	Medical	36	M	Yes	No	Mixed	Agent_full	Yes	DOC	Individual
5 Chamberland et al. 2015a	Medical	53	Low	Yes	Yes	WSE	Agent_part	Yes	DOC	Individual
6 Chamberland et al. 2015b	Medical	54	Low	Yes	Yes	Mixed	Agent_part	No	DOC	Individual
7 Eva et al. 2007 I	Medical	60	Low	Yes	Yes	During	Agent_part	No	DOC	Individual
8 Eva et al. 2007 II	Medical	48	Low	Yes	Yes	During	Agent_part	No	DOC	Individual
9 Gold et al. 2013	Teacher	120	Low	Yes	Yes	Mixed	Agent_part	Yes	DOC	Individual
10 Gutiérrez-Maldonado et al. 2014	Medical	42	Low	Yes	No	After	Agent_full	Yes	INT	Collaboration
11 Heitzmann et al. 2013	Medical	152	High	No	Yes	WSE	No	No	DOC	Individual
12 Heitzmann et al. 2015	Medical	98	High	No	Yes	WSE	No	No	DOC	Individual
13 Heitzmann et al. 2018a	Teacher	108	High	No	Yes	WSE	No	No	DOC	Individual
14 Hellermann et al. 2015	Teacher	96	Low	Yes	Yes	Mixed	Agent_part	Yes	M	Collaboration
15 Ibiapina et al. 2014 I	Medical	58	WSE	Yes	No	During	Agent_full	Yes	DOC	Individual
16 Ibiapina et al. 2014 II	Medical	57	WSE	Yes	No	During	Agent_full	Yes	DOC	Individual
17 Jarodzka et al. 2012	Medical	60	Low	No	Yes	No	No	No	INT	Individual
18 Klug et al. 2016	Teacher	47	High	Yes	No	M	Agent_full	Yes	INT	Individual
19 Krammer et al. 2016	Teacher	129	Low	Yes	No	Mixed	Agent_full	Yes	DOC	Collaboration
20 Liaw et al. 2010	Medical	30	Low	Yes	Yes	Mixed	No	Yes	INT	Collaboration
21 Mamede et al. 2012	Medical	46	High	Yes	No	Mixed	Agent_full	Yes	DOC	Individual
22 Mamede et al. 2014	Medical	110	High	Yes	No	Mixed	Agent_full	Yes	DOC	Individual
23 Neisstadt and Smith 1997	Medical	127	High	Yes	No	Longterm	Agent_full	No	DOC	Individual
24 Ohst et al. 2015 I	Teacher	45	Low	Yes	Yes	During	Agent_part	No	INT	Individual
25 Ohst et al. 2015 II	Teacher	56	Low	Yes	Yes	During	Agent_part	No	INT	Individual
26 Papa et al. 2007	Medical	123	Low	No	Yes	No	Agent_part	No	DOC	Individual

Table 2 Summary of the primary studies

Authors, year	Participants	Measurements	Learning environment	Summary of results
1 Baghdady et al. 2014	2nd year medical students (dental medicine)	Written: diagnostic accuracy of dental abnormalities	Individual computer-based learning and testing environment	The inclusion of the basic science test improved performance on a test of diagnostic accuracy
2 Bahreini et al. 2013	In-service nurses	Expert rating of performance on the Nurse Competence Scale	University-based teaching hospital, learning in groups	The professional portfolios help nurses update their knowledge, skills, and competence
3 Besser et al. 2015	In-service teachers in Mathematics	Written: pedagogical content knowledge & knowledge about formative assessment	Further education center; group training, individual implementation	Teachers trained in formative assessment of teaching competency-oriented mathematics outperformed teachers in general training on the expertise test
4 Chamberland et al. 2011	3rd year medical students	Expert rating of clinical reasoning: diagnostic accuracy and performance	Individual work on clinical cases in university settings	Self-explanation seems to be an effective technique to help medical students learn clinical reasoning
5 Chamberland et al. 2015a	3rd year medical students	Expert rating of clinical reasoning: diagnostic accuracy and performance	Individual work on clinical cases in university settings	Self-explanation & examples seems to be an effective technique to help medical students learn clinical reasoning
6 Chamberland et al. 2015b	3rd year medical students	Expert rating of clinical reasoning: diagnostic accuracy and performance	Group work on clinical cases in university settings	Self-explanation & examples seems to be an effective technique to help medical students learn clinical reasoning
7 Eva et al. 2007 I	Undergraduate psychology students	Clinical reasoning in diagnosing cardiac conditions via ECG	Individual work on clinical cases in university settings	The instruction to utilize a combined reasoning approach resulted in greater diagnostic accuracy
8 Eva et al. 2007 II	Undergraduate psychology students	Clinical reasoning in diagnosing cardiac conditions via ECG	Group work on clinical cases in university settings	The instruction to utilize a combined reasoning approach resulted in greater diagnostic accuracy
9 Gold et al. 2013	Preservice teachers	Performance on video clips assessment, according to developed scale	Group work on cases in university settings	Positive effect of guided video analysis on professional vision of classroom management behavior
10 Gutierrez-Maldonado et al., 2014	Undergraduate psychology students	Performance: interviewing and differential diagnosis skills	Individual work, simulated interviews with virtual patients	Group trained in formative assessment performed better in diagnosing mental disorders in virtual patients
11 Heitzmann, et al., 2013	3rd and 4th year nursing students	Expert rating: strategic and conditional knowledge + paper test for declarative knowledge	Individual work in computer-based learning environment with worked-examples	Self-explanation prompts for erroneous worked examples had a negative effect on strategic and conditional knowledge
12 Heitzmann et al. 2015	3rd and 4th year medical students	Expert rating: strategic and conditional knowledge + paper test for declarative knowledge	Individual work in computer-based learning environment with worked-examples	Adaptable feedback had a positive effect on diagnostic competence, whereas self-explanations did not further improved them

Table 2 (continued)

Authors, year	Participants	Measurements	Learning environment	Summary of results
13 Heitzmann et al. 2018a	Preservice teachers	Diagnostic competences (strategic, conditional, and declarative knowledge)	Individual work in computer-based learning environment with worked examples	Error-explanation prompts had a positive effect on declarative-conceptual knowledge if combined with adaptable feedback, but a negative on learning of practical knowledge
14 Hellemann et al. 2015	3rd year teacher students	Video-based test for classroom management knowledge + situational judgment test for strategic knowledge	Examples of classroom management, analysis of video individually and in groups	Analyzing others and one's own teaching was the most efficient training program for improving professional vision
15 Ibiapina et al. 2014 I	5th year medical students	Expert rating: diagnostic accuracy-based responses to clinical cases	Individual work on clinical cases in university settings	When practicing the diagnosing of cases, exemplary and cued reflections were more beneficial for learning than free reflection
16 Ibiapina et al. 2014 II	6th year medical students	Expert rating: diagnostic accuracy based responses to clinical cases	Individual work on clinical cases in university settings	When practicing the diagnosing of cases, exemplary and cued reflections were more beneficial for learning than free reflection
17 Jarodzka et al. 2012	Final year medical students	MCQ for clinical reasoning and virtual observations of patients, and interpretation	Modeling: expert model demonstrates how to visually search and interpret symptoms	Spotlight EVMC can successfully convey clinical reasoning based on visual observations
18 Klug et al. 2016	In-service teachers	Scenario test to measure diagnostic competences, MCQ for knowledge in diagnostics, + professional self-concept test	Group work on cases during session, filling in reflection diaries	Training program enhanced teachers' diagnostic competence, especially actions before and during diagnosing
19 Krammer et al., 2016	1st year teacher students	IRT scale: agreement with experts in analyzing video cases	Group work on analysis of video cases (own or others' teaching)	Positive effect of video-based training for developing professional vision and related knowledge
20 Liaw et al. 2010	1st year nursing students	Checklist developed by researcher: clinical performance of students	Group work: completing simulation and discussing case scenario	Incorporation of simulation learning activities into problem-based discussion was an effective strategy for teaching nursing students to assess and manage crisis events
21 Mamede et al. 2012	4th year medical students	Expert rating: accuracy in diagnosing cases (diagnostic accuracy score)	Individual work: participants diagnosed 6 clinical cases under different exp. conditions	Structured reflection while practicing with cases appears to foster the learning of clinical knowledge
22 Mamede et al. 2014	4th year medical students	Expert rating: accuracy in diagnosing cases (diagnostic accuracy score)	Individual work: participants diagnosed	Structured reflection while practicing with cases enhanced learning of diagnosis both of the diseases practiced and of their alternative diagnoses

Table 2 (continued)

Authors, year	Participants	Measurements	Learning environment	Summary of results
23 Neistadt and Smith 1997	1st semester senior undergraduate students in occupational therapy	Expert rating: diagnostic reasoning skills coded from treatment plans assignments	4 clinical cases under different exp. conditions Learning and assignments based on videos (occupational therapy evaluations of real patients with neurological dysfunction)	Treatment group outperformed the control group in identifying problems of video 1, but not video 2. The classroom-as-clinic methodology needs to be combined with explicit coaching to be effective
24 Ohst et al. 2015 I	Preservice teachers	Expert rating: identification test (open-ended) and a speed test on learning strategies + quality of reasoning	Computer-based learning environment to learn how to classify learning strategies as well as how to diagnose them	While achieving approximately equal learning outcomes, the treatment group needed less learning time and stated higher interest
25 Ohst et al. 2015 II	Undergraduate teacher students	Expert rating: different types of knowledge in a definition task, identification and explanation tasks, and verification tasks	Computer-based learning environment on how to classify primary learning strategies	The treatment group (categorical framework during pretraining) claimed to possess higher levels of interest, achieved higher learning outcomes, and learned more efficiently
26 Papa et al. 2007	2nd year medical students	Computer-based knowledge test: accuracy in diagnosing clinical cases	Individual work: computer-based tutor to enhance differential diagnosis skills	The ability to perform differential diagnosis is enhanced by training based upon principles of cognitive sciences
27 Peixoto et al. 2017	4th year medical students	Expert rating: accuracy in diagnosing clinical cases + adherence to the procedure	Individual work: modeling use of self-explanation prompts in analyzing data for diagnosis	Self-explanation of pathophysiology did not improve students' diagnostic performance for all diseases, but only for ones similar to the example
28 Raupach et al. 2010 I	4th year medical students	Summative written assessments of ECG interpretation skills, important to pass the course	Small-group, near-peer teaching during a 6-week cardiorespiratory course	Assessment format appeared to be more powerful than choice of instructional method in enhancing student learning. The effect observed in the second cohort was masked by an overriding incentive generated by the summative assessment in the first cohort
29 Raupach et al. 2010 II	4th year medical students	Rather formative written assessment of ECG interpretation skills	Small-group, near-peer teaching during a 6-week cardiorespiratory course	

Table 2 (continued)

Authors, year	Participants	Measurements	Learning environment	Summary of results
30 Raupach et al. 2013 I	4th year medical students	Correct identification of ECG diagnoses	Learning in different formats: self-directed learning or small-group peer teaching facilitated by more advanced students	Summative assessments increased the odds of correctly identifying ECG diagnoses, of spending extra time on ECG self-study and of using additional learning material
31 Raupach et al. 2013 II	4th year medical students	Correct identification of ECG diagnoses	Learning in different formats: self-directed learning or small-group peer teaching facilitated by more advanced students	Summative assessments increased the odds of correctly identifying ECG diagnoses, of spending extra time on ECG self-study and of using additional learning material
32 Round 1999	4th year medical students	Written: diagnostic thinking inventory to measure clinical reasoning	Group work on a clinical scenario + discussing clinical algorithm	Students participating in the teaching intervention were significantly better on the diagnostic thinking inventory than controls
33 Stark et al. 2011 I	3-5th year medical students	Expert rating + knowledge test: specific conceptual, strategic and conditional knowledge	Group work with cases and worked examples: four learning conditions (varied presence of errors and type of feedback)	Erroneous examples were effective in combination with elaborated feedback with respect to strategic and conditional knowledge
34 Stark et al. 2011 II	3-5th year medical students	Expert rating + knowledge test: specific conceptual, strategic and conditional knowledge	Group work with cases and worked examples: four learning conditions (varied presence of errors and type of feedback)	Erroneous examples were effective in combination with elaborated feedback with respect to conceptual, strategic, and conditional knowledge
35 Sunder et al. 2016	2nd year teacher students	Professional vision of learners measured through analysis of video-cases (compared to expert solution)	Group work: theoretical explanations, analyses of classroom situations of other teachers (text or video) + own micro-teaching followed by analysis of it	Both presentation formats (text and video) had significant positive effects on learning to apply professional vision framework to lesson analysis

Table 3 Effects of professional knowledge base, context factors, and instructional support on diagnostic competences

Summary effect (random model)	<i>p</i> of <i>Q</i>	<i>Q</i> (<i>p</i>)	<i>g</i>	95% CI	<i>n</i> (<i>k</i>)	<i>tau</i> ²	<i>I</i> ²	<i>EV</i> (<i>N</i>)	<i>R</i> -Index
1. Treatment vs. Control	<.001	166.65	0.39	[0.22; 0.56]	3472 (35)	0.18	79.60	1 (12)	0.51
<i>Moderator variables</i>	Significance of moderator	<i>Q</i> between	Effect Size	95% CI	<i>n</i> (<i>k</i>)	Heterogeneity	<i>I</i> ²	Quality assessment	<i>R</i> -Index
2. Professional knowledge base	<i>p</i> of <i>Q</i>	<i>Q</i> between	<i>g</i>	95% CI		<i>tau</i> ²		<i>EV</i> (<i>N</i>)	
Low	<.01	6.98	0.48	[0.25; 0.71]	1161 (16)	0.14	64.58	1 (5)	0.56
High			0.27	[0.05; 0.50]	2141 (17)	0.16	82.04	1 (5)	0.48
3. Instructional Support									
Problem-solving included	<.001	19.09	0.51	[0.34; 0.67]	2008 (25)	0.10	58.40	1 (10)	0.54
Problem-solving not included			0.20	[-0.21; 0.60]	930 (8)	0.30	88.68	1 (2)	0.47
Reflections included	<.001	17.11	0.58	[0.36; 0.80]	1432 (18)	0.15	66.93	1 (8)	0.54
Reflections not included			0.26	[0.01; 0.50]	1506 (15)	0.18	80.06	1 (4)	0.48
Examples included	.06, ns	2.85	0.33	[0.09; 0.58]	1704 (20)	0.24	67.37	1 (6)	0.54
Examples not included			0.39	[0.20; 0.58]	1768 (17)	0.10	58.40	1 (6)	0.52
Roles included	<.001	35.79	0.49	[0.20; 0.58]	2472 (26)	0.17	74.65	1 (10)	0.58
Roles not included			0.00	[-0.18; 0.17]	1214 (9)	0.04	54.14	0 (2)	0.36
Prompts included	<.05	5.33	0.47	[0.09; 0.58]	2003 (26)	0.12	63.44	1 (8)	0.52
Prompts not included			0.26	[-0.01; 0.54]	1547 (12)	0.18	80.06	1 (4)	0.46
<i>Type of Prompts</i>	.071, ns	5.03							
Prompts after			0.63	[0.11; 1.15]	409 (4)	0.19	70.36	1 (2)	0.44
Prompts during			0.36	[0.11; 0.61]	606 (8)	0.06	48.51	1 (2)	0.50
Prompts longterm			1.05	[0.08; 2.02]	267 (3)	0.66	90.57	1 (2)	0.82
Prompts mixed			0.38	[0.16; 0.60]	785 (11)	0.11	32.68	0 (4)	0.48
4. Knowledge base*Scaffolding									
Problem-solving	.34, ns	0.69							
Low knowledge base			0.41	[0.23; 0.59]	866 (13)	0.03	23.35	0 (4)	0.56
High knowledge base			0.59	[0.26; 0.91]	1027 (10)	0.20	76.88	1 (4)	0.54
<i>Examples</i>	<.001	22.87							
Low knowledge base			0.52	[0.25; 0.79]	839 (12)	0.14	64.02	1 (4)	0.57
High knowledge base			0.18	[-0.31; 0.67]	647 (6)	0.33	88.45	1 (2)	0.52
<i>Roles</i>	<.05	4.04							
Low knowledge base			0.39	[0.20; 0.57]	836 (12)	0.02	22.95	0 (3)	0.60
High knowledge base			0.45	[0.16; 0.73]	1561 (12)	0.20	84.56	1 (4)	0.52
Prompts	.19, ns	1.22							

Table 3 (continued)

<i>Summary effect (random model)</i>	<i>p of Q</i>	<i>Q (p)</i>	<i>g</i>	<i>95% CI</i>	<i>n(k)</i>	<i>tau^2</i>	<i>I^2</i>	<i>EV(N)</i>	<i>R-Index</i>
Low knowledge base			0.39	[0.20; 0.57]	812(12)	0.03	22.55	0(3)	0.56
High knowledge base			0.61	[0.22; 1.01]	841(9)	0.28	79.31	1(3)	0.54
Reflection phases	.14, ns	1.63							
Low knowledge base			0.45	[0.18; 0.72]	603(8)	0.07	46.29	0(3)	0.57
High knowledge base			0.67	[0.22; 1.13]	714(8)	0.34	80.21	1(3)	0.55
5. Context Factors									
5.1 Domain									
Medicine	<.01	6.01	0.23	[0.15; 0.51]	2764(26)	0.15	78.27	1(9)	0.47
Teacher Education			0.58	[0.17; 0.98]	708(9)	0.31	82.46	1(3)	0.67
5.2 Diagnostic Situation									
Interaction Based	<.001	23.58	0.77	[0.40; 1.13]	578(10)	0.25	71.65	1(5)	0.61
Document Based			0.27	[0.10; 0.44]	2922(26)	0.13	73.64	1(7)	0.48
5.3 Processing mode									
Collaboration	.52, ns	0.40	0.36	[0.12; 0.61]	1428(11)	0.12	73.12	1(3)	0.52
Individual			0.36	[0.15; 0.56]	2258(26)	0.21	80.06	1(9)	0.51

n number of participants, *k* number of studies, *EV*evidential value; 1 achieved; 0 flat line, *N* number of significant *p* values

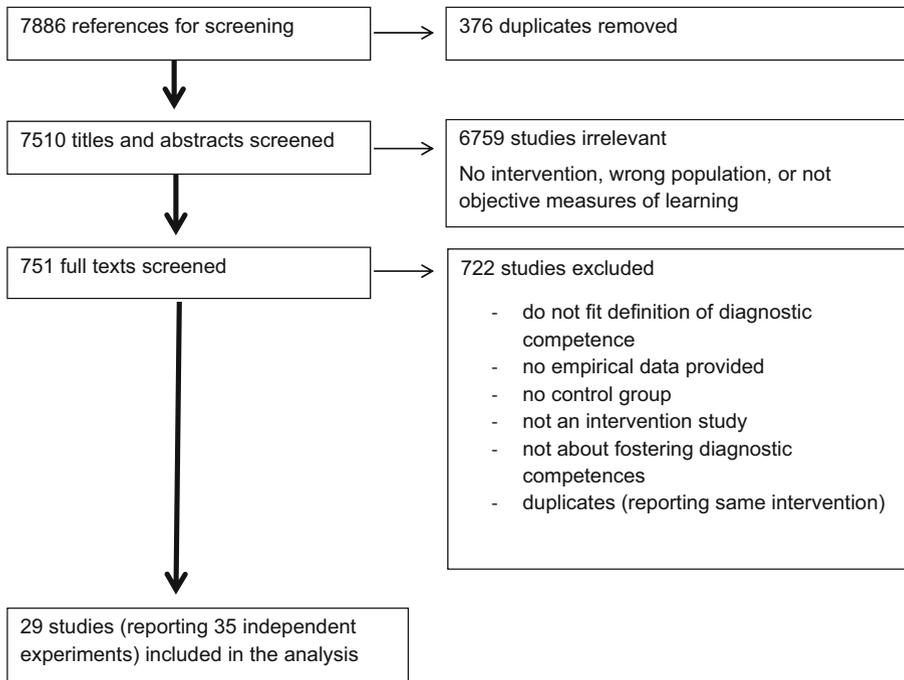


Fig. 1 Flow chart of the study selection process

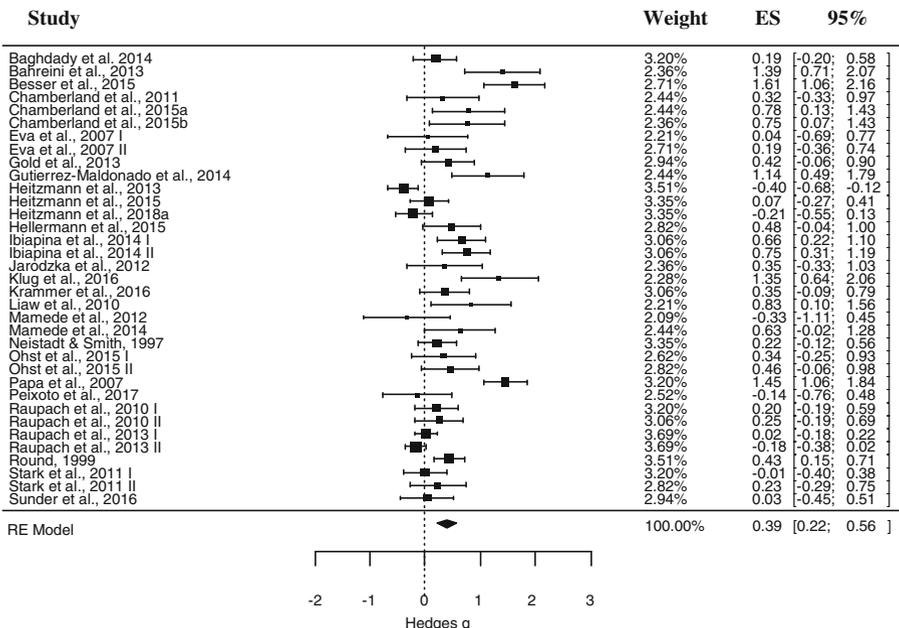


Fig. 2 Forest plot of the overall effect of instruction on diagnostic competences

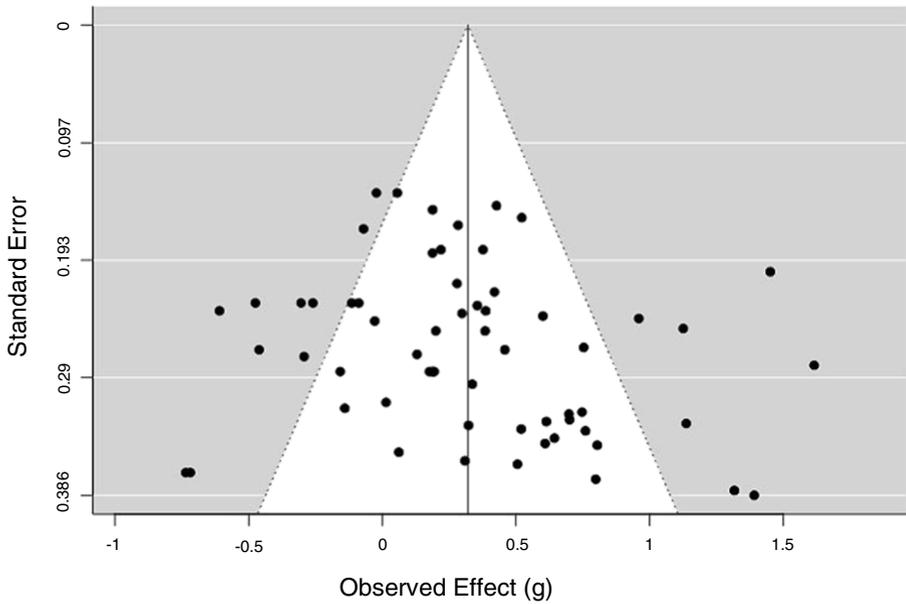


Fig. 3 Funnel plot of the overall effect of instruction on diagnostic competences

Diagnostic situation included the (a) *information base*, which was coded as “interaction-based” if information to diagnose was gathered through real or simulated interactions; as “document-based” if information was gathered from a document or a video with diagnostic information without a possibility to interact; and (b) *processing mode*, which was coded as “collaboration” if collaboration during the diagnostic process was necessary, or as “individual” if diagnosis/decision was achieved by a single individual.

Diagnostic competences were coded as outcome measures. The measure of diagnostic competence was coded as “procedural” if application of knowledge in solving diagnostic cases

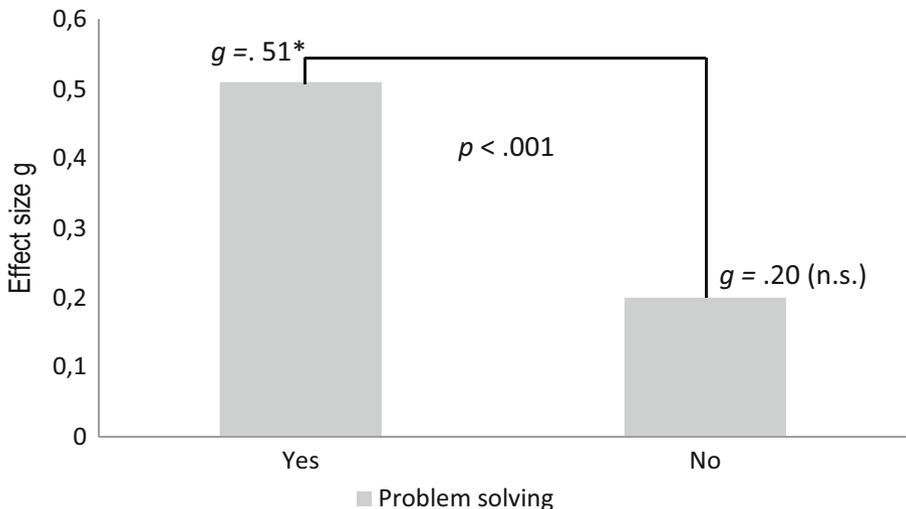


Fig. 4 The effect of problem-solving (elements included vs. not included during learning phases)

was measured at the posttest. It was coded as “strategic” if application of knowledge to a specific case was not required, but rather the knowledge was measured on a conceptual level concerning the strategy of the diagnostic process (e.g., learners were asked for the diagnostic steps without applying their knowledge to a case). Diagnostic competences were coded as “conceptual” if diagnostic steps or processes were measured on the conceptual level and concerned terms or interrelations of terms (as an example in medicine, knowledge of liver disease); in teaching, it could refer to understanding learning processes.

Studies reporting comparison of two levels of the same moderator (i.e., comparing students with high versus low levels of prior professional knowledge or comparing prompts during versus after the diagnosis processes) within one study were coded as within-study effects (“WSE”) and were excluded from respective moderator analyses.

Statistical Methods

Calculation of the Effect Sizes and Synthesis of the Analysis

This meta-analysis used a random-effects model and adjusted effect size estimation (Hedges g). To address the often-complicated design of the studies, we employed multiple comparisons and therefore correlated meta-regression samples using robust variance estimation (Tanner-Smith et al. 2016). However, robust-variance estimation models are intended neither to provide precise variance parameter estimates nor to test the null hypotheses regarding heterogeneity (Tanner-Smith et al. 2016). To overcome these limitations, we used additional meta-analytic procedures recommended for subgroup analysis (Borenstein et al. 2009). To get a

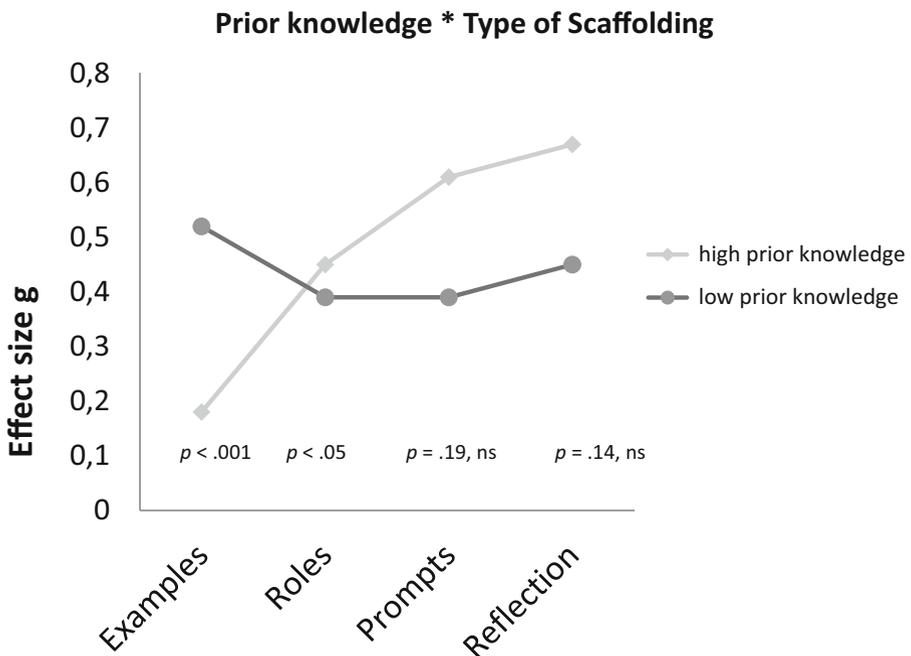


Fig. 5 Interaction between levels of prior professional knowledge and scaffolding

representative result, we used only one effect size per study in the moderator analysis. If the study reported multiple effects, a small-scaled meta-analysis was run to synthesize the results within a single study before including the effect in the summary effect estimation. Also, we used confidence intervals to assess the significance of an effect. We used multiple heterogeneity estimates (Q -statistics, τ^2 , I^2) to determine the variance of the true effect sizes between studies, its statistical significance, and the proportion of this variance that can be explained by random factors. In addition, we used the thresholds suggested by Higgins et al. (2003) to interpret the I^2 : 25% for low heterogeneity, 50% for medium heterogeneity, and 75% for high heterogeneity.

Assessment of Publication Bias and Questionable Research Practices

This meta-analysis on the effects of instructional support on facilitating diagnostic competences includes primary studies from medical and teacher education and combines studies with large and relatively small samples. To address these issues, we used a range of statistical methods to control and correct for possible publication bias, questionable research practices, and other manipulations to ensure sufficient power, validity, and generalizability of the findings. Because the approaches used to detect and estimate publication bias have different assumptions and limitations, we used a combination of those methods with the assumption that if a strong indicator either for or against publication bias occurred, the results of all methods applied to test for publication bias point in the same direction.

The first approach is based on a graphical representation of the relationship between effect sizes and the standard error. Egger's test in the absence of publication bias and questionable research practices assumes that studies are evenly distributed on both sides of the average, but if publication bias is present, reported effect sizes correlate with sample sizes (Sterne and Egger 2001). Trim'n'fill techniques can be used to correct for any identified asymmetry (Duval and Tweedie 2000). The weakness of the funnel plot-based methods is that they do not take true heterogeneity into account and cannot distinguish between methodological-caused biases and true differences between study effects.

The second approach, the p -curve analysis, addresses both detection and correction for possible publication bias and evaluates the significance of estimated effect sizes (Simonsohn

Table 4 Interaction between domain and instructional support moderator variables

Factor (yes)	Medical education			Difference	Teacher education		
	<i>k</i>	<i>SE</i>	<i>g</i>		<i>g</i>	<i>SE</i>	<i>k</i>
Problem-solving	16	0.11	0.50**	ns	0.51**	0.12	9
Examples	4	0.39	0.68	ns	0.42*	0.10	7
Prompts	20	0.10	0.46**	ns	0.72**	0.18	4
Roles	16	0.09	0.45**	ns	0.60**	0.18	8
Reflection	12	0.13	0.53**	ns	0.68**	0.24	6
Interaction-based	4	0.22	0.93**	ns	0.67**	0.27	6
Document based	22	0.10	0.27*	ns	0.25	0.20	4
Individual	22	0.11	0.35*	ns	0.44	0.31	4
Collaborative	6	0.11	0.20	**	0.57	0.25	5

k number of studies

* $p < .05$; ** $p < .01$

et al. 2015). This technique provides a robust estimate of the significance of p values from the studies, plots them, and combines the half and full p curve to make inferences about an evidential value; however, it is based only on significant p values.

The third approach, which takes under consideration both significant and insignificant results, is the R-index (R-Index.org. 2014). It can be used to examine the credibility and replicability of studies. The R-index can be between 0 and 100% (Schimmack 2016); values below 22% indicate the absence of a true effect; values below 50% indicate inadequate statistical power of the study; and values above 50% are acceptable to support credibility and replicability of the results, although values above 80% are preferred.

Results

Results of the Literature Search

The search resulted in 7510 articles after deleting duplicates. During abstract and full-text screening, most excluded studies were non-empirical, had no control group, or had measuring outcomes that did not fit the definition of diagnostic competences (Fig. 1). The 35 eligible studies (published between 1997 and 2018) provided 60 effect size estimations. The studies and their characteristics are presented in Table 1 in alphabetical order. The total sample consisted of 3472 participants. Most of the studies (69%) implemented random assignment to control and experimental condition. The sample of studies provided almost equal distribution of participants, with low (16 studies) and high (17 studies) professional knowledge bases.

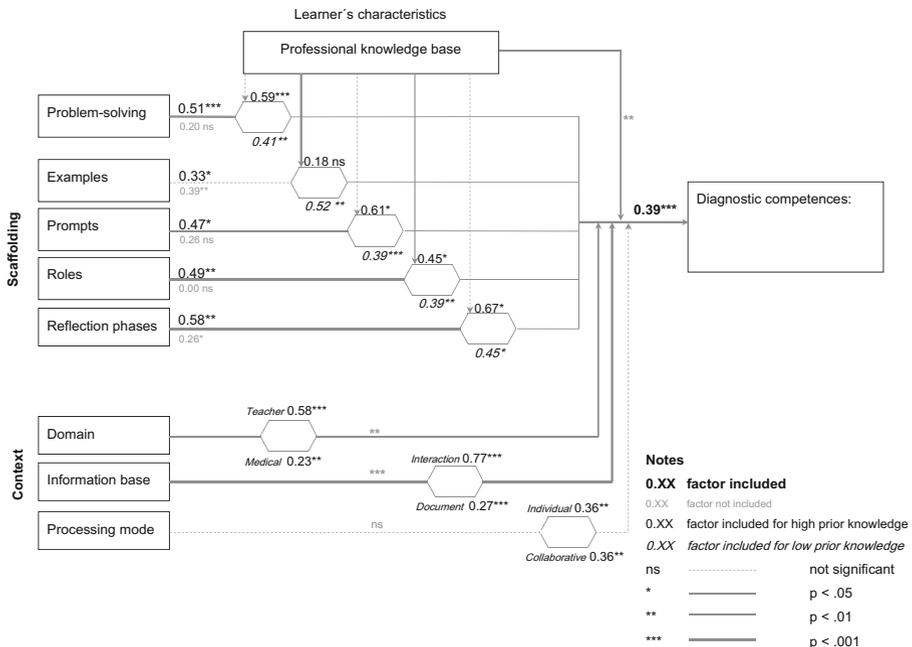


Fig. 6 Fostering diagnostic competences

Some moderator levels included in the coding scheme were not present in the sample of studies. Specifically, all studies measured and reported a procedural aspect of diagnostic competences; however, none of the primary studies reported assessment of conceptual knowledge gain or the strategic aspect of diagnostic competences separately from procedural aspect. Additionally, regarding role-taking, 26 studies (74%) reported assigning roles during learning; all of them reported assigning an agent role, either for the whole learning process (53%) or for parts of it (42%), and two more studies reported multiple conditions.

Among the 35 studies, 25 included problem-solving (71%), 8 studies did not include problem-solving (23%), and 2 reported within-study effects. Only one study reported no use of any type of scaffolding, and instead used only explicit presentation of information to facilitate the advancement of competences. All (100%) of the 25 studies that included problem-solving had at least one type of additional scaffolding.

Results of Quality and Preliminary Analysis

The procedures targeted at assessing the quality of primary studies and the generalizability of the summary and moderator effects found in the meta-analysis indicated no evidence of publication bias or questionable research practices. The Eggers test for funnel plot asymmetry was insignificant ($z = 1.58$; $p = 0.11$). However, the p -curve analysis indicated that six results out of 38 provided insufficient evidential value. Furthermore, the R-index analyses indicated that 10 results out of 38 have inadequate replicability indexes. These findings limit the generalizability of evidence for research question 4, due to insufficient data from the primary studies (Table 3).

The meta-regression on control variables (year of publication, publication type, lab, design of study) showed that these factors do not explain a statistically significant amount of variance between study effects (p values above .05).

Summary Effect of the Instructional Support on the Diagnostic Competences

Regarding research question 1, instructional support was found to have a medium positive effect ($g = .39$; $p = .001$; 95% CI [.22; .56]), on fostering diagnostic competences in the combined sample of studies in medical and teacher education. The effect has sufficient evidential value and an acceptable replicability index. The analysis also identified high heterogeneity between studies ($\tau^2 = .18$; $I^2 = 79.60\%$), justifying further moderator analyses. The effect sizes found in individual studies, weights, and confidence intervals, as well as the summary effect from the random effect model estimation, are presented in alphabetical order in Fig. 2. A funnel plot of effect size distribution and standard errors is presented in Fig. 3.

Effects of Moderators

Effect of the Professional Knowledge Base

Regarding research question 2, subgroup analyses (Table 3) indicate that learners with a lower level of prior professional knowledge showed a higher increase in diagnostic competences ($g = .48$, $SE = .12$, $p < .05$) than learners with a higher level of prior professional knowledge, whose diagnostic competences also increased through instructional intervention ($g = .27$, $SE = .11$, $p < .05$).

Effect of Problem-Solving

Regarding research question 3, the studies in the meta-analysis provided evidence in favor of learning through problem-solving (Fig. 4) as the instructional approach to enhance diagnostic competences. Including problem-solving elements in instruction ($g = .51$, $SE = .11$, $p < .05$) was more beneficial than not including problem-solving ($g = .20$, $SE = .11$, $p = ns$) for advancing diagnostic competences. The studies provided sufficient evidential value. The moderator role of problem-solving instructions was statistically significant ($Q(1, 31) = 19.09$, $p < .001$).

Effect of Scaffolding

Despite descriptive differences (Table 3), settings including examples compared with settings not including examples did not reach statistical significance regarding effects on advancement of diagnostic competences ($Q(1, 35) = 2.85$, $p = .06$).

Role-taking (taking an agent's role during the learning phase) had a significant positive effect on advancing diagnostic competences ($g = .49$, $SE = .11$, $p < .05$). Primary studies with settings in which roles were not assigned during learning indicated no statistically significant effect on the advancement of diagnostic competences ($g = 0$, $SE = .09$, $p > .05$). Assigning roles was a statistically significant moderator ($Q(1, 33) = 19.09$, $p < .001$).

Including prompts had a significantly higher positive effect on diagnostic competences ($g = .47$, $SE = .09$, $p < .05$) than not including prompts ($g = .26$, $SE = .14$, $p < .05$); the moderator was significant as well ($Q(1, 37) = 5.33$, $p < .05$). More specifically, the types of prompts coded in relation to the diagnosis were presented during, after, long-term, or a mixture of these. The type of prompts as moderator did not reach statistical significance ($Q(3, 22) = 5.03$, $p = .071$). Note, however, that providing prompts *after* the diagnosis tended to be more beneficial for the learners than providing prompts *during* diagnostic processes or combining multiple types of prompts. Providing long-term prompts also tended to be beneficial for advancing diagnostic competences (Table 3).

Reflection phases had a significantly higher positive effect on advancing diagnostic competences ($g = .58$, $SE = .11$, $p < .05$) compared to instructional support not including reflection phases ($g = .26$, $SE = .11$, $p < .05$). This moderator was statistically significant ($Q(1, 31) = 17.11$, $p < .001$).

Interaction Between Professional Knowledge Base and Instructional Support

Regarding research question 4, problem-solving was identified as effective for learners with high ($g = .59$, $SE = .17$, $p < .05$) and low ($g = .41$, $SE = .09$, $p < .05$) levels of prior professional knowledge. If problem-solving was not included ($k = 8$), there was no statistically significant gain in competence for learners with a high level of prior professional knowledge ($g = -.10$, $SE = .12$, $p > .05$; $k = 5$), nor for learners with a low level of prior professional knowledge ($g = .67$, $SE = .45$, $p > .05$; $k = 3$). In interpreting these findings, the relatively low number of primary studies that were used in the analysis must be considered (Table 3).

As hypothesized, more advanced learners benefited most from types of scaffolding that afforded higher levels of self-regulation, namely, reflection phases: ($g = .67$, $SE = .23$, $p < .05$). Providing examples instead of problem-solving activities to learners with a high level of prior professional knowledge did not lead to advancement of their diagnostic competences ($g = .18$, $SE = .25$, $p > .05$). In contrast to advanced learners, learners with a low level of prior professional knowledge benefited from examples ($g = .52$, $SE = .14$, $p < .05$). These findings support

the hypothesis regarding the degree of self-regulated problem-solving for learners with low vs. high levels of professional knowledge (Fig. 5). However, other measures of instructional support, such as prompts, had similar positive effects on the advancement of diagnostic competences for learners with low as well as high levels of prior professional knowledge. These results did not contribute sufficiently to evaluating the hypothesis, as data from primary studies provided an insufficient evidential value (Table 3).

Effect of Contextual Factors

Regarding research question 5, the diagnostic situation significantly moderated the effects of instructional support on the advancement of diagnostic competences ($Q(1, 35) = 23.58, p < .01$). The diagnostic competences were significantly more advanced through interaction-based activities ($g = .77, SE = .19, p < .05$) than through document-based activities ($g = .27, SE = .08, p < .05$). The necessity for collaboration (processing mode) failed to reach significance as a moderator ($Q(1, 36) = 0.40, p = .52$).

Regarding medical and teacher education, the statistical analysis showed significant variance between the subgroups ($Q(1, 33) = 6.01, p < .05$). However, insignificant results of meta-regression indicate that the “domain” was not a statistically significant moderator to explain the differences found. Thus, the differences in the magnitude of the effects for medical ($g = .33, SE = .09, p < .05, n = 26$) and teacher ($g = .58, SE = .21, p < .05, n = 9$) education are likely due to an unequal amount of studies representing the two fields.

To address the possible difference between the domains, we conducted a post hoc analysis to estimate whether the *average* effect sizes for the moderators of prior professional knowledge, instructional support measures, and contextual factors differ significantly for medical education and teacher education.

Problem-solving was included in more than half of the studies in medical education ($N = 16$) and all studies in teacher education ($N = 9$), resulting in the same average effects with similar standard errors (Table 4). The positive effect of including examples had achieved significance only in medical education; however, no significant differences between the domains were found. Assigning roles, providing prompts, and reflection phases had similar positive effects in both domains; the differences in the magnitude of the effects were not significant, which might be due to an unequal amount of studies in the moderator levels.

The analysis indicated that there was a significant difference in prior professional knowledge between the two domains; moreover, there was evidence of the interaction between levels of prior knowledge and the domain affecting the development of diagnostic competences. The mean effect for high prior professional knowledge in teacher education was significantly greater ($g = .88, SE = .64, p > .05, n = 3$) than the one in medical education ($g = .15, SE = .09, p > .05, n = 14$); however, neither of the effects individually reached statistical significance. In contrast, the mean effect for low prior knowledge in medical education ($g = .57, SE = .19, p < .01, n = 10$) was significantly higher than that for teacher education ($g = .34, SE = .10, p < .05, n = 6$); both effects individually were positive and significant. The only further difference in contextual factors was for the collaborative processing mode (Table 3). Learners in teacher education ($g = .57, SE = .25, p > .05, n = 5$) benefited significantly more from collaboration than did learners in medical education ($g = .20, SE = .11, p > .05, n = 6$); however, neither of the effects reached statistical significance individually. Therefore, we suggest that this pattern of findings indicates initial evidence to support the claim that the findings concerning instructional support can be generalized across the two domains.

Summary and Conclusion

This meta-analysis (see Fig. 6 for the overview) shows that interventions for facilitating diagnostic competences in the investigated domains of medical education and teacher education are particularly effective if they involve learners in some form of problem-solving. Advancing diagnostic competences without the learners' own engagement in problem-solving seems unlikely. This is true for both the low and high levels of professional knowledge investigated. Most studies that addressed forms of problem-centered instructional support additionally provided one or several types of scaffolding (see Hmelo-Silver et al. 2007). Approaches to scaffolding that come in addition to learners' own problem-solving—namely, assigning roles, providing prompts, and reflection phases—have clear positive effects on diagnostic competences. Overall, this analysis shows no indication for different effect sizes of different types of scaffolding. However, with respect to the timing, prompts tend to be more effective if they are provided several times over a longer period of time or after the learner's own problem-solving activity compared to prompts delivered during the problem-solving activity itself. The effectiveness of scaffolding approaches depends on the learners' prior professional knowledge base. Reflection phases are more effective for more advanced learners, whereas providing examples is effective for less advanced learners.

Theoretical Significance

The findings of this meta-analysis have implications that support the claim that interventions on problem-solving skills necessarily need to involve the learners in problem-solving activities (Anderson 1983; Van Lehn 1996), and that this claim can be generalized to the area of solving complex medical- and teaching-related diagnostic problems. Moreover, the meta-analysis yielded evidence in support of generalizing the medium-sized positive overall effects of scaffolding found in studies in other fields (Kim et al. 2018) to medical and teacher education. Additionally, the findings suggest generalization of the well-known positive effect of examples for novice learners rather than for the more advanced learners. Moreover, beyond these generalizations of what has already been established in different fields, the findings of the meta-analysis also contribute to an advancement of the scientific understanding of scaffolding.

First, the expertise reversal effect—that is, a negative effect of scaffolding for more advanced learners (see Kalyuga et al. 2003)—could not be established for the studies we reviewed in medical and teacher education. On the contrary, most of the scaffolding types we reviewed yielded positive effects for learners with a greater knowledge base as well.

Second, this meta-analysis contributes to a better understanding of how and why different types of scaffolds may cause different effects. The initially derived hypothesis on the interaction of scaffolding type and prior knowledge received partial support through the analyses. Indeed, learners with more advanced prior knowledge benefited more from scaffolds affording their self-regulated diagnostic problem-solving activity. However, rather than a continuous dimension with increasing degrees of freedom, scaffolding appears to show a dichotomous distinction: as long as learners are able to practice problem-solving, all of the different types of scaffolding are beneficial.

When more advanced learners are hindered in their problem-solving activity through scaffolding, enforcing alternative activities interventions remains largely ineffective. This explanation might seem to be related to the expertise reversal effect (Kalyuga et al. 2003);

however, it is not the same. The expertise reversal effect would assume negative effects of all types of unnecessary scaffolding. Rather, the findings of this meta-analysis show positive effects of different types of scaffolding for more advanced learners, suggesting that this group is able to make good use of the support. This finding seems rather supportive of the so-called Sesame Street or Matthew effect (Walberg and Tsai 1983), indicating that learners with better prerequisites are also better in exploiting offerings originally designed to support learners with less well-developed prerequisites. We can speculate that rather than the degree of freedom for self-regulated activity, it is the fidelity of the problem-solving activity that determines the effects of scaffolding on the learning of more advanced learners.

If the scaffolding changes the learners' activities away from the diagnostic problem-solving process, then learners with more advanced knowledge would suffer from an expertise reversal effect. If the scaffolding leaves the targeted problem-solving processes untouched but supports the learners to productively engage in them, then a Sesame Street effect is likely to happen. It seems worthwhile to build on this initial explanation in developing a more theory-based classification of different types of scaffolding. Types of scaffolds may differ in how much self-directed problem-solving they afford and require from learners. Scaffolding is thus not just something for beginners. However, we currently know little about the processes through which more advanced learners benefit from scaffolding. This meta-analysis cannot contribute to this issue beyond pointing to the need for more primary studies that include more advanced learners, scaffolding, and process analyses.

Third, the findings of this meta-analysis advance our scientific understanding of other aspects of scaffolding, using prompts, at least for diagnostic problem-solving but probably beyond. Prior meta-analyses have shown a limited overall effect of scaffolding through prompts (Kim et al. 2018). Our study addresses the timing of prompts. With respect to advancing the procedural aspects of diagnostic competences, prompts during diagnostic problem-solving seem less effective than more long-term prompts or prompts that help in understanding the diagnostic processes after the engagement in them. It seems plausible to assume that prompts in the ongoing problem-solving process are meant to avoid failures of learners' problem-solving through in-process guidance, whereas prompts after the diagnostic problem-solving are typically meant to afford guidance and stimulate reflection. This is in line with models on learning through problem-solving that emphasize how important it is for expertise development that learners self-regulate when they identify and correct errors in their knowledge base (Kapur and Rummel 2012). Therefore, this finding on the timing of prompts may generalize to other types of complex skill development through problem-solving.

Finally, this meta-analysis contributes to our understanding of the generality of the effects of instructional support measures in medical and teacher education. Diagnosing is a goal-oriented collection of information to make decisions in both domains, and scaffolding has quite similar patterns of effects in advancing diagnosis in medical and teacher education. This is not a trivial finding. For example, a recent meta-analysis has shown that scaffolding can have quite different effects across different domains (Kim et al. 2018). Comparable effects on outcomes may be taken as initial evidence that instructional support measures can be transferred between the domains. A meta-analysis cannot deliver evidence that the processes of learning are also comparable without conducting more primary studies focusing on the learning process. The effects found on levels of prior knowledge in medical and teacher education might be explained by the fact that the structure of professional knowledge development in medical education seems to be traced better (i.e., the curriculum introduces topics and practice opportunities in stable order), whereas in teacher education, the students

themselves have more control over the sequence and even the amount of topics they engage in during their studies. It is thus more difficult in teacher education to infer prior professional knowledge based on the semester of study (Linninger et al. 2015).

Limitations

The limitations to the generalizability of findings from this meta-analysis are due primarily to insufficient data and the complex experimental designs from the primary studies. First, too few studies address diagnostic competences to do further in-depth comparisons of the two domains of teacher and medical education. The studies from different contexts were analyzed together without yet having the statistical power to look for domain and context effects of the instructional support.

Second, the four scaffolding categories presented in this meta-analysis unite the scaffolding types that can further be divided into subcategories according to theoretical background. It would have been favorable to distinguish each of the scaffolds further, that is, to distinguish between reflections upon the problem at hand and reflections upon own diagnostic reasoning, or between different kinds of prompts, such as providing additional information, self-explanation prompts, and meta-cognitive prompts. However, due to the low number of studies, such comparisons are not presently possible in a meta-analytical way. Furthermore, most of the primary studies included a combination of different types of examples, of reflection phases, and of prompts. For example, some of the studies include combinations of meta-cognitive prompts and prompts providing additional information on the problem. Other studies combine, for example, reflection on the diagnostic situation with reflection on the own reasoning. Therefore, even though reflection or prompts seemed to have positive effects overall, different kinds of reflections or prompts were subsumed, and thus differential effects of the different kinds of reflections or prompts are still possible. Therefore, conclusions about the effectiveness of each scaffolding type should be made with caution; however, the presented analysis offers insight into how scaffolding types might be categorized according to the self-regulation required. If more empirical studies with detailed descriptions of used scaffolding are not available, further systematic analysis with more precise categorization will contribute to explaining more heterogeneity; this analysis was not performed in this study.

Third, the use of multiple instructional support measures in a large proportion of the primary studies precluded direct comparisons of the effectiveness of scaffolding measures and evaluating the effects of scaffolding on different components of diagnostic competences.

Fourth, most of the studies used a combination of tests for assessing learning and reported combined competence measures and global ratings; therefore, the effects of instructional support on different types of professional knowledge (conceptual, practical, or strategic) were not estimated.

The *p*-curve analysis indicated that the studies in the analysis do not provide sufficient evidential value for some levels of moderators. This is true for not assigning roles to learners during the intervention, providing different types of prompts simultaneously, and exploring the effects of scaffolding on learners with low levels of prior knowledge. Therefore, the corresponding statistically significant results have to be cautiously interpreted (see Table 3). Furthermore, replicability indexes vary considerably for studies within the different levels of the moderators (0.36–0.82). The values below 0.50 indicate inadequate statistical power of the effect; however, generalization of these effects is limited and requires primarily more empirical studies addressing specific roles of different scaffolding procedures and learning outcomes.

Recommendations for Practice

Diagnostic competences may develop with increasing experience in practice. However, a thoroughly planned higher education program would care for practicing opportunities to start this process much earlier during the program. Evidence from this meta-analysis shows that interventions that include problem-solving activities of the learners have the potential to advance the procedural aspects of diagnostic competences. This fact has been recognized in medical education for many years (e.g., Vernon and Blake 1993). Additionally, results of this meta-analysis show that the potential advancement of diagnostic competences through problem-solving interventions is at least as large if not larger in teacher education than in medical education. Traditional lectures and courses with examples but without students' own problem-solving activities may be good in developing the necessary conceptual and strategic knowledge base, but these teaching formats will probably not contribute much to advancing the procedural aspects of diagnostic competences. Additional instructional guidance through scaffolding is likely to further improve learning through problem-solving. At least for the procedural aspects of diagnostic competences, prompts are more promising if delivered after, rather than during, the problem-solving process. When solving diagnostic problems, learners with little prior professional knowledge are likely to benefit when solving diagnostic problems from additional examples more than from other types of scaffolding. More advanced learners still benefit from scaffolding, but they gain more from types of scaffolding that afford their self-regulated problem-solving.

Further Research

The findings of the current research synthesis provide some insights about differences and similarities in the fields of medical and teacher education and enhance the scientific understanding of the role of instruction, context, and prior professional knowledge base in the facilitation of diagnostic competences. The study also identified several further questions to be addressed by experimental studies and further research syntheses. For example, more primary studies with a design that allows for direct comparisons of scaffolding types are needed to further validate the model suggesting the placement of scaffolding measures on a continuum from high levels of guidance to more self-regulation and meta-cognition. Moreover, more primary studies are needed that report not only global scores, but components of these scores and more specific descriptions of learning and testing activities. Those studies would enable addressing the effects of different types of instruction and scaffolding on components of diagnostic competences (conceptual, procedural, strategic knowledge, analytical and decision-making skills, and epistemic-diagnostic activities).

Another promising direction would be to focus on creating, validating, and implementing scales for the assessment of diagnostic competence to address different components of diagnostic competences within and across domains. More standardized measures such as these would also support identifying what types of scaffolding optimally meet the needs of learners with different levels of prior professional knowledge.

Additionally, further research may also explore the motivational aspects of diagnostic competences more systematically by including subjective measures of learning outcomes (e.g., perceived utility, confidence in applying learned strategies, or self-perceived competence levels).

Acknowledgements The research for this article was funded by the German Research Association (Deutsche Forschungsgemeinschaft, DFG) (FOR2385).

Appendix 1 Coding manual

Table 5 Coding manual: study characteristics and statistics (low inference items)

Study characteristic	Description
Study ID	Identification number for each primary study
Author(s)	Individual(s) who published a particular primary study
Year	Publication year of an article or manuscript write up
Publication type	Type of publication (journal article, dissertation, etc.)
Study design	Random distribution of treatment and control condition: code 1 if authors explicitly mention that randomization took place while assigning participants into treatment and control groups. code 0—no randomization took place.
ES ID	Identification number for each effect size within each study (if one study reports multiple effect sizes)
Total sample size	Total number of individuals who participated in the experiment
Sample size (sub-groups)	Number of participants in each treatment or experimental group, and control condition or group, as reported by authors of the study
Statistical data	Descriptive statistics (means, standard deviations, standard errors, or confidence intervals) for a particular outcome measure as it is reported in the results section of a primary study.
Participants	Description of the participants provided by authors (usually includes domain of study and year of education)
Participants' mean age	Age of individuals who participated in the experiment

Table 6 Coding manual: prior knowledge (high inference item)

Main category	Description	Code	Example
Professional knowledge	<i>Global estimate</i> of learner's experiences based on <i>study year</i> or on direct statement of <i>familiarity of the learning context</i> in the manuscript.	Low	For example, in the study by Liaw et al. (2010), first year nursing undergraduate students in a bachelor degree program participated in the study. Hence prior knowledge—low. The cutoff depends on the length of the study program as well as on if the study is a postgraduate program. That is, in a 4-year program, second year is considered low. In a 2-year program, second year is considered high. Postgraduate programs are usually considered high. Authors mention that topic or context was unfamiliar to students. For example: Peixoto et al. (2017) authors mention that the 4th year students “had little clinical experience and it was expected that they have not yet developed and /or consolidated rich illness scripts of the diseases used in the study.”
		High	Authors mention that the students were in the final year of their education (i.e., internal medicine rotation) or already have in-service experience in medicine or teaching. Authors mention that the topic or context is familiar to students. For example: Heitzmann et al. (2015): “Intervention was implemented using computer-supported environment (CACUS), which all students were familiar to and worked during their regular studies.”

Table 7 Coding manual: instructional support measures (during the training phase of the experiments)

Main category	Subcategory	Code	Description
Instructional approach	Problem-solving	Yes	Yes: learners receive cases (scenarios, vignettes, etc.) and make diagnoses themselves during the learning phase.
		No	
Scaffolding	Examples	Yes	Yes: learners receive examples (modeling, worked out examples or similar) at some time during the training; examples do not have to be present throughout the whole training phase.
		No	
	Prompts (overall)	Yes	Yes: learners receive hints of how to handle the learning material in a certain way in order to support the learner anytime during or after the diagnosis.
		No	
	Prompts (during)	Yes	Yes: prompts were given to the learner during task completion. These typically include self-explanation prompts, meta-cognitive prompts, and cognitive tutors. It is important that the coded support measure takes place during task completion (i.e., before diagnosing) and focuses either on fostering elaboration, reasoning, or meta-cognitive processes.
		No	
	Prompts (after)	Yes	Yes: prompts which support the learner after the diagnosis has been completed (i.e., a diagnosis has been given!).
	Prompts (long-term)	Yes	Yes: prompts to focus on various steps of the diagnostic process that take place over a long period of time. This measure typically includes learning diaries/journals or multiple sessions with a phase used for reflecting on experiences and is part of longitudinal studies.
		No	
	Role during the training phase	AGR (agent)	AGR
OR (other)			OR: learners were in the role of the observer or a person with the characteristic or process to be diagnosed (patient, respectively student)
No			No: participants were not assigned to any of the roles (i.e., during lectures or in example-based learning)
Reflection phases		Yes	Yes: authors explicitly mention “reflection phases” or any phases during learning (or treatment) when learners were encouraged to think about the goals of the procedures they complete, analyze their own performance, and plan further steps.
		No	

Table 8 Coding manual for diagnostic situation

Main category	Subcategory	Code	Description
Nature of the diagnostic situation	Information base	INT	INT: information to diagnose is gathered through real or simulated interactions.
		DOCB	DOCB: information to diagnose is gathered from a document or a video that contains diagnostic information without the possibility of interaction
Necessity for collaboration Domain	Processing mode	COLL	COLL: collaboration during the diagnostic process is necessary
		IND	IND: diagnosis is achieved by a single individual.
		TE	TE: teacher education
		MED	MED: medical education

Table 9 Coding manual for diagnostic competences

Main category	Subcategory	Code	Description	Example
Diagnostic competence	Conceptual outcomes	Yes	Yes: outcome measure on conceptual knowledge: as knowledge on terms and on interrelations in a specific take out of reality	In medicine, knowledge of liver disease/heart disease. For teaching, it could be the understanding of learning processes.
		No	Yes: during the posttest, learners are solving diagnostic cases. The application of knowledge is measured.	
	Procedural outcomes	Yes	Yes: the diagnostic process is measured on a conceptual level.	Learners solve diagnostic cases as assessment or perform a procedure they learned.
		No		
	Strategic outcomes	Yes		For example, learners are asked for the diagnostic steps without applying their knowledge to a case.
		No		

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

*Primary studies in the meta-analysis

- Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine*, *68*(1), 52–81 PMID: 8447896.
- Anderson, J. R. (1983). *Cognitive science series. The architecture of cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>.
- *Baghdady, M., Carnahan, H., Lam, E. W. N., & Woods, N. N. (2014). Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical Education*, *48*(2), 181–188. <https://doi.org/10.1111/medu.12302>.
- *Bahreini, M., Moattari, M., Shahamat, S., Dobaradaran, S., & Ravanipour, M. (2013). Improvement of Iranian nurses' competence through professional portfolio: A quasi-experimental study. *Nursing & Health Sciences*, *15*(1), 51–57. <https://doi.org/10.1111/j.1442-2018.2012.00733.x>.
- Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, *1996*(68), 3–12. <https://doi.org/10.1002/tl.37219966804>.
- Beauchamp, C. (2015). Reflection in teacher education: Issues emerging from a review of current literature. *Reflective Practice: International and Multidisciplinary Perspectives*, *16*(1), 123–141. <https://doi.org/10.1080/14623943.2014.982525>.
- Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 505–518). Dordrecht: Springer. https://doi.org/10.1007/978-1-4614-3185-5_39.
- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, *87*(2), 309–344. <https://doi.org/10.3102/0034654316670999>.
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577. <https://doi.org/10.1016/j.learninstruc.2007.09.007>.
- *Besser, M., Leiss, D., & Klieme, E. (2015). Wirkung von Lehrerfortbildungen auf Expertise von Lehrkräften zu formativem Assessment im kompetenzorientierten Mathematikunterricht. *Zeitschrift für*

- Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 110–122. <https://doi.org/10.1026/0049-8637/a000128>.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Brush, T., & Saye, J. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning*, 1(2), 1–12 Retrieved from: <http://www.ncolr.org/jiol/issues/pdf/1.2.3.pdf>.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. <https://doi.org/10.2307/1170684>.
- *Chamberland, M., St-Onge, C., Setrakian, J., Lanthier, L., Bergeron, L., Bourget, A., Mamede S., Schmidt H., Rikers R. (2011). The influence of medical students' self-explanations on diagnostic performance. *Medical Education*, 45(7), 688–695. <https://doi.org/10.1111/j.1365-2923.2011.03933.x>.
- *Chamberland, M., Mamede, S., St-Onge, C., Setrakian, J., & Schmidt, H. G. (2015a). Does medical students' diagnostic performance improve by observing examples of self-explanation provided by peers or experts? *Advances in Health Sciences Education: Theory and Practice*, 20(4), 981–993. <https://doi.org/10.1007/s10459-014-9576-7>.
- *Chamberland, M., Mamede, S., St-Onge, C., Setrakian, J., Bergeron, L., & Schmidt, H. (2015b). Self-explanation in learning clinical reasoning: The added value of examples and prompts. *Medical Education*, 49(2), 193–202. <https://doi.org/10.1111/medu.12623>.
- Charlin, B., Tardif, J., & Boshuizen, H. P. A. (2000). Script and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine: Journal of the Association of American Medical Colleges*, 75(2), 182–190.
- Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education*, 48(8), 750–760. <https://doi.org/10.1111/medu.12473>.
- Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, 28(6), 557–573. <https://doi.org/10.1111/j.1365-2729.2011.00476.x>.
- Dewey, J. (1933). *How we think. A restatement of the relation of reflective thinking to the educative process (revised ed.)*. Boston: D. C. Heath.
- Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, 13(5), 533–568. [https://doi.org/10.1016/S0959-4752\(02\)00025-7](https://doi.org/10.1016/S0959-4752(02)00025-7).
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- *Eva, K. W., Hatala, R. M., Leblanc, V. R., & Brooks, L. R. (2007). Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Medical Education*, 41(12), 1152–1158. <https://doi.org/10.1111/j.1365-2923.2007.02923.x>.
- Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1), 56–66. <https://doi.org/10.1080/00461520.2012.748005>.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(2), 28–45. <https://doi.org/10.14786/flr.v2i2.96>.
- Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., Möller, G. E., Wiesbeck, A., & Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: Effects of e-learning with video cases and role-play. *Instructional Science*, 43(4), 443–462. <https://doi.org/10.1007/s11251-014-9341-6>.
- Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology*, 45(6), 1097–1114. <https://doi.org/10.1111/bjet.12188>.
- Glogger-Frey, I., Herppich, S., & Seidel, T. (2018). Linking teachers' professional knowledge and teachers' actions: Judgment processes, judgments and training. *Teaching and Teacher Education*, 76(1), 176–180. <https://doi.org/10.1016/j.tate.2018.08.00>.
- *Gold, B., Förster, S., Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27(3), 141–155. <https://doi.org/10.1024/1010-0652/a000100>.
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055–2100 Retrieved from: <http://www.tcrecord.org/15018>.

- *Gutiérrez-Maldonado, J., Ferrer-García, M., Pla-Sanjuanelo, J., & Andres-Pueyo, A. (2014). Virtual humans and formative assessment to train diagnostic skills in bulimia nervosa. *Studies in Health Technology and Informatics*, 199(1), 30–34. <https://doi.org/10.3233/978-1-61499-401-5-30>.
- Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (pp. 115–140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- *Heitzmann, N., Fischer, F., & Fischer, M. R. (2013). Förderung von Diagnosekompetenz bei Lehrern: Differenzierte Effekte von Selbsterklärungsprompts. In Vortrag auf der 78. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) der DGfE, Dortmund, September 25–27, 2013.
- *Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, 49(10), 993–1003. <https://doi.org/10.1111/medu.12778>.
- *Heitzmann, N., Fischer, F., & Fischer, M. R. (2018a). Worked examples with errors: When self-explanation prompts hinder learning of teachers diagnostic competences on problem-based learning. *Instructional Science*, 46(2), 245–271. <https://doi.org/10.1007/s11251-017-9432-2>.
- Heitzmann, N., Timothy, V., & Fischer, F. (2018b). Förderung von Diagnosekompetenzen durch Simulationen: Zwei Metaanalysen. In Presented at GEBF Conference (15–17 February 2018). Basel.
- *Hellermann, C., Gold, B., & Holodynski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium: Die Wirkung der Analyse eigener und fremder Unterrichts Videos auf das strategische Wissen und die professionelle Wahrnehmung. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 47(2), 97–109. <https://doi.org/10.1026/0049-8637/a000129>.
- Helmke, A., Schrader, F.-W., & Helmke, T. (2012). EMU: Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung. Unterrichtsdiagnostik – Ein Weg, um Unterrichten sichtbar zu machen. *Schulverwaltung Bayern*, 35(6), 180–183.
- Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, 31(5), 3–15. <https://doi.org/10.3102/0013189X031005003>.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107. <https://doi.org/10.1080/00461520701263368>.
- *Ibiapina, C., Mamede, S., Moura, A., Elói-Santos, S., & van Gog, T. (2014). Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Medical Education*, 48(8), 796–805. <https://doi.org/10.1111/medu.12435>.
- *Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40(5), 813–827. <https://doi.org/10.1007/s11251-012-9218-5>.
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem solving learning outcomes. *Educational Technology Research & Development*, 45(1), 45–94. <https://doi.org/10.1007/BF02299613>.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4.
- Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science*, 40(4), 645–650. <https://doi.org/10.1007/s11251-012-9235-4>.
- Kaufman, D., Yoskowitz, A., & Patel, V. L. (2008). Clinical reasoning and biomedical knowledge: Implications for teaching. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (pp. 137–150). Amsterdam: Elsevier Health Sciences.
- Kiesewetter, J., Ebersbach, R., Görlitz, A., Holzer, M., Fischer, M. R., & Schmidmaier, R. (2013). Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions. *PLoS One*, 8(8), e71486. <https://doi.org/10.1371/journal.pone.0071486>.
- Kim, N. J., Belland, B. R., & Walker. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review*, 30(2), 397–429. <https://doi.org/10.1007/s10648-017-9419-1>.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1.
- *Klug, J., Gerich, M., & Schmitz, B. (2016). Can teachers' diagnostic competence be fostered through training and the use of a diary? *Journal for Educational Research Online*, 8(3), 184–206. urn:nbn:de:0111-pedocs-128256.

- Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, 6(1), 3–34. <https://doi.org/10.1007/BF00155578>.
- *Krammer, K., Frommelt, M., Furrer Auf der Maur, G., Biaggi, S., Hugener, I., (2016). Videos in der Ausbildung von Lehrkräften: Förderung der professionellen Unterrichtswahrnehmung durch die Analyse von eigenen bzw. fremden Videos. *Unterrichtswissenschaft*, 44 (4), 357–372. <https://doi.org/10.3262/UW1604357>.
- *Liaw, S. Y., Chen, F. G., Klainin, P., Brammer, J., O'Brien, A., & Samarasekera, D. D. (2010). Developing clinical competency in crisis event management: An integrated simulation problem-based learning activity. *Advances in Health Sciences Education: Theory and Practice*, 15(3), 403–413. <https://doi.org/10.1007/s10459-009-9208-9>.
- Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D., Seidel, T., Terhart, E., & Kunter, M. (2015). Assessing teachers' educational knowledge: Construct specification and validation using mixed methods. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(2), 62–74. <https://doi.org/10.1026/0049-8637/a000126>.
- Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education*, 3(1), 15–25. <https://doi.org/10.1016/j.hpe.2017.01.003>.
- *Mamede, S., van Gog, T., Moura, A. S., de Faria, R. M. D., Peixoto, J. M., Rikers, R. M. J. P., & Schmidt, H. G. (2012). Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical Education*, 46(5), 464–472. <https://doi.org/10.1111/j.1365-2923.2012.04217.x>.
- *Mamede, S., van Gog, T., Sampaio, A. M., de Faria, R. M. D., Maria, J. P., & Schmidt, H. G. (2014). How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Academic Medicine*, 89(1), 121–127. <https://doi.org/10.1097/ACM.0000000000000076>.
- *Neistadt, M. E., & Smith, R. E. (1997). Teaching diagnostic reasoning: Using a classroom-as-clinic methodology with videotapes. *American Journal of Occupational Therapy*, 51(5), 360–368. <https://doi.org/10.5014/ajot.51.5.360>
- Nguyen, Q. D., Fernandez, N., Karsenti, T., & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education*, 48(12), 1176–1189. <https://doi.org/10.1111/medu.12583>.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.12691/education-3-1-8>.
- O'Connell, T. S., & Dymont, J. E. (2011). The case of reflective journals: Is the jury still out? *Reflective Practice*, 12(1), 47–59. <https://doi.org/10.1080/14623943.2011.541093>.
- *Ohst, A., Glogger, I., Nückles, M., & Renkl, A. (2015). Helping preservice teachers with inaccurate and fragmentary prior knowledge to acquire conceptual understanding of psychological principles. *Psychology Learning & Teaching*, 14(1), 5–25. <https://doi.org/10.1177/1475725714564925>
- *Papa, F. J., Oglesby, M. W., Aldrich, D. G., Schaller, F., & Cipher, D. J. (2007). Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. *Medical Education*, 41(4), 419–425. <https://doi.org/10.1111/j.1365-2929.2006.02693.x>.
- *Peixoto, J. M., Mamede, S., de Faria, R. M. D., de Moura, A. S., Santos, S. M. E., & Schmidt, H. G. (2017). The effect of self-explanation of pathophysiological mechanisms of diseases on medical students' diagnostic performance. *Advances in Health Sciences Education* 22(5), 1183–1197. <https://doi.org/10.1007/s10459-017-9757-2>.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, 13(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4.
- *Raupach, T. N., Raupach, T., Hanneforth, N., Anders, S., Pukrop, T., Th. J. Ten Cate O, & Harendza, S. (2010). Impact of teaching and assessment format on electrocardiogram interpretation skills. *Medical Education*, 44(7), 731–740. <https://doi.org/10.1111/j.1365-2923.2010.03687.x>
- *Raupach, T., Brown, J., Anders, S., Hasenfuss, G., & Harendza, S. (2013). Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *BMC Medical Education*, 11(3), 61–71. <https://doi.org/10.1186/1741-7015-11-61>
- Reiser, B. J., & Tabak, I. (2014). Scaffolding. In R. K. Sawyer (Ed.), *Cambridge handbooks in psychology: The Cambridge handbook of the learning sciences* (pp. 44–62). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.005>.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3.

- Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, 31(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3.
- R-Index.org (2014). *R-index 1.0*. www.r-Index.org.
- *Round, A. P. (1999). Teaching clinical reasoning – A preliminary controlled study. *Medical Education*, 33(7), 480–483. <https://doi.org/10.1046/j.1365-2923.1999.00352.x>
- Sandars, J. (2009). The use of reflection in medical education: AMEE Guide No. 44. *Medical Teacher*, 31(8), 685–695. <https://doi.org/10.1080/01421590903050374>.
- Schimmack, U. (2016). *The replicability-index: Quantifying statistical research integrity*. Retrieved from. <https://replicationindex.com/2016/01/31/a-revised-introduction-to-the-r-index/>
- Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, 34(1), 56–65. <https://doi.org/10.1016/j.tate.2013.03.004>.
- Shulman, L. S. (1987). Knowledge and teaching. Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>.
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud and ambitious P-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, 144(6), 1146–1152. <https://doi.org/10.1037/xge0000104>.
- *Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction*, 21(1), 22–33. <https://doi.org/10.1016/j.learninstruc.2009.10.001>.
- Stegmann, K., Pilz, F., Siebeck, M., & Fischer, F. (2012). Vicarious learning during simulations: Is it more effective than hands-on training? *Medical Education*, 46(10), 1001–1008. <https://doi.org/10.1111/j.1365-2923.2012.04344.x>.
- Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046–1055. [https://doi.org/10.1016/S0895-4356\(01\)00377-8](https://doi.org/10.1016/S0895-4356(01)00377-8).
- Stoof, A., Martens, R. L., van Merriënboer, J. G., & Bastiaens, T. J. (2002). The boundary approach of competence: A constructivist aid for understanding and using the concept of competence. *Human Resource Development Review*, 1(3), 345–365. <https://doi.org/10.1177/1534484302013005>.
- Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior*, 26(4), 491–494. <https://doi.org/10.1016/j.chb.2009.08.006>.
- Stürmer, K., Seidel, T., & Holzberger, D. (2016). Intra-individual differences in developing professional vision: Preservice teachers' changes in the course of an innovative teacher education program. *Instructional Science*, 44(3), 293–309. <https://doi.org/10.1007/s11251-016-9373-1>.
- *Sunder, C., Todorova, M., & Möller, K. (2016). Förderung der professionellen Wahrnehmung bei Bachelorstudierenden durch Fallanalysen. Lohnt sich der Einsatz von Videos bei der Repräsentation der Fälle? *Unterrichtswissenschaft*, 44(4), 339–356. <https://doi.org/10.3262/UW1604339>.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5).
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.
- Tabak, I., & Kyza, E. (2018). Research on scaffolding in the learning sciences: A methodological perspective. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences, 191–200*. NY: Routledge.
- Tanner-Smith, E., Tipton, E., & Polanin, J. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial. *Journal of Developmental and Life-Course Criminology*, 2(1), 85–112. <https://doi.org/10.1007/s40865-016-0026-5>.
- Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., et al. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical Teacher*, 34(6), 421–444. <https://doi.org/10.3109/0142159X.2012.68>.
- Trempler, K., Hetmanek, A., Wecker, C., Kiesewetter, J., Wermelt, M., Fischer, F., Fischer, M., & Gräsel, C. (2015). Nutzung von Evidenz im Bildungsbereich. Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und Bewertung von Studien. *Zeitschrift Für Pädagogik*, 61, 143–166.
- Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology*, 12(3), 323–355. <https://doi.org/10.1080/09541440050114543>.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- Van Lehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47(1), 513–539. <https://doi.org/10.1146/annurev.psych.47.1.513>.

- Vernon, D., & Blake, R. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine*, 68(7), 550–563.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511. <https://doi.org/10.1007/s10648-016-9361-7>.
- Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind and society: The development of higher psychological processes* (pp. 79–91). Cambridge, MA: Harvard University Press.
- Walberg, H. J., & Tsai, S.-I. (1983). Matthew effects in education. *American Educational Research Journal*, 20(3), 359–373. <https://doi.org/10.2307/1162605>.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>.
- Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology*, 94(4), 660–668. <https://doi.org/10.1037/0022-0663.94.4.660>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.