



Toward an Overarching Framework for Systematizing Research Perspectives on Diagnostic Thinking and Practice

Timo Leuders · Katharina Loibl · Daniel Sommerhoff ·
Stephanie Herppich · Anna-Katharina Praetorius

Received: 2 March 2021 / Accepted: 18 February 2022 / Published online: 15 March 2022
© The Author(s) 2022

Abstract Assessing students' learning processes and products is considered a core requirement of effective teaching. As such, it is an object of research in several disciplines and research areas. To structure the various corresponding research perspectives and provide a broader, yet still systematic view of the field, we propose an overarching framework that allows for systematizing foci of interest, goals, methodologies, and theoretical premises as four integral aspects of conducting research in this area. We demonstrate the benefits of the overarching framework by using it as a tool to analyze and systematize previous research from four different research perspectives. Based on this framework, we discuss the strengths and limitations of existing studies and, in particular, highlight theoretical premises that are rarely explicitly addressed but become more obvious by using the overarching framework. In addition, we provide directions for future research by drawing connections within and between research perspectives. Our analyses illustrate how the overarching framework can provide a foundation for research synthesis and inform future studies.

Keywords Diagnostic thinking · Diagnostic practice · Research synthesis · Research map

Timo Leuders (✉) · Katharina Loibl
University of Education Freiburg, Freiburg, Germany
E-Mail: leuders@ph-freiburg.de

Daniel Sommerhoff
Department for Mathematics Education, Leibniz Institute for Science and Mathematics Education,
Kiel, Germany

Stephanie Herppich
DLR Project Management Agency, Bonn, Germany

Anna-Katharina Praetorius
University of Zurich, Zurich, Switzerland

Auf dem Weg zu einem übergreifenden Rahmen für die Systematisierung von Forschungsperspektiven zu diagnostischem Denken und Handeln

Zusammenfassung Die Bewertung der Lernprozesse und -produkte von Schülerinnen und Schülern gilt als eine zentrale Voraussetzung für effektiven Unterricht. Als solche ist sie Gegenstand der Forschung in verschiedenen Disziplinen und Forschungsbereichen. Um die verschiedenen Forschungsperspektiven zu strukturieren und eine breitere, aber dennoch systematische Sicht auf das Feld zu ermöglichen, schlagen wir einen übergreifenden Rahmen vor, der es ermöglicht, Interessensschwerpunkte, Ziele, Methoden und theoretische Prämissen als vier integrale Aspekte der Forschung in diesem Bereich zu systematisieren. Um die Vorteile des übergreifenden Rahmens aufzuzeigen, nutzen wir ihn als Instrument zur Analyse und Systematisierung bisheriger Forschung aus vier verschiedenen Forschungsperspektiven. Auf Grundlage dieses Rahmens diskutieren wir die Stärken und Grenzen bestehender Studien und heben insbesondere theoretische Prämissen hervor, die selten explizit dargestellt werden, aber durch die Verwendung des übergreifenden Rahmens deutlicher werden. Darüber hinaus geben wir Hinweise für künftige Forschung durch das Aufzeigen von Verbindungen innerhalb und zwischen verschiedenen Forschungsperspektiven. Unsere Analysen veranschaulichen, wie der übergreifende Rahmen eine Grundlage für Forschungssynthesen bilden und künftige Forschung anregen kann.

Schlüsselwörter Diagnostisches Denken · Diagnostisches Handeln · Forschungssynthese · Forschungslandkarte

1 Introduction: Teachers' Diagnostic Thinking and Practice

Assessing students' learning processes and products is considered a core requirement for effective teaching (Helmke 2010; Herppich et al. 2018; Leuders et al. 2017; Schrader 2011). Although this statement is undisputed, there are considerably diverse notions on what exactly assessment is focused on and how it is enacted, depending on the research perspective taken.

The term “assessment” is sometimes used to simply denote instruments and procedures to assess students (e.g., Deluca et al. 2016). However, like many researchers (e.g., Bennett 2011; Messick 1995; Van den Heuvel-Panhuizen 1996), we adopt a broader and integrative view: Assessment includes the development, application, and evaluation of tests or other instruments as well as any classroom practice and interpretive evaluation of student thinking during learning with the goal of procuring information on students' traits and states, learning processes, and learning outcomes in a specific context. In this sense, we use the broader term of “teachers' diagnostic thinking and practice” in the following instead of assessment.

Diagnostic thinking is usually understood as the processing of information related to assessment (Leuders et al. 2017; Schrader and Praetorius 2018), while diagnostic practices range from formative probing in the context of on-the-fly adaptation of

teaching to summative testing in the context of grading (cf. Herppich et al. 2018; Shavelson et al. 2008). It can be geared, among others, to the assessment of mathematical knowledge and skills, mathematical ways of thinking and working, or the interests of learners.

Teachers' diagnostic thinking and practice are studied in several disciplines and research areas, leading to a broad literature base on this topic. Attempting to summarize and systematize the field on an international level, scholars have initiated several collaborative projects over the last years. Each project focused on specific aspects of the field. The "NeDiKo" group, for example, provided a conceptual model that frames diagnostic competence (considered as a bundle of constructs comprising stable and measurable knowledge, skills, motivations, and beliefs connected to diagnostic practice)¹ and a process model that describes teachers' diagnostic thinking and practice (Herppich et al. 2018). The "Cosima" framework focuses on supporting the acquisition of expertise with respect to diagnostic practice via simulations during teacher training (Heitzmann et al. 2019). The "DiaCoM" framework focuses on the theoretical description and empirical investigation of cognitive processes that constitute diagnostic judgments (Loibl et al. 2020). Although each framework provides an overview for its specific focus, an overarching framework is missing that allows to comprehensively systematize the field.

Individual researchers could locate their individual projects within this overarching framework and in doing so, be aware of their own perspective within a "bigger picture" to avoid losing sight of complementary perspectives. The community could benefit as well from such an overarching framework, as it allows for more cumulative work in the field, based on a shared understanding of terms and structure (see also Grossman and McDonald 2008). Additionally, relating existing research to such an overarching framework would allow identifying the blind spots of different research projects and under-investigated areas in the field.

As a step toward an overarching framework for research on teachers' diagnostic thinking and practice, this article and the presented framework pursue the aims

- A. to structure research in the field of teachers' diagnostic thinking and practice with regard to (1) foci of interest, (2) goals, (3) methodologies, and (4) theoretical premises of different research perspectives;
- B. to identify the strengths and limitations of previous research with a structured analysis and to provide orientation to inform future research.

To achieve these aims, we first introduce the overarching framework (Sect. 2), subsequently use it to analyze examples of research perspectives on teachers' diagnostic thinking and practice regarding their strengths and limitations (Sect. 3), and, finally, identify directions for future research (Sect. 4).

¹ Internationally, the terms competence and competency are used. In this article, we use competence, as it is more common in the European context. For the difference between the terms see Blömeke et al. (2015).

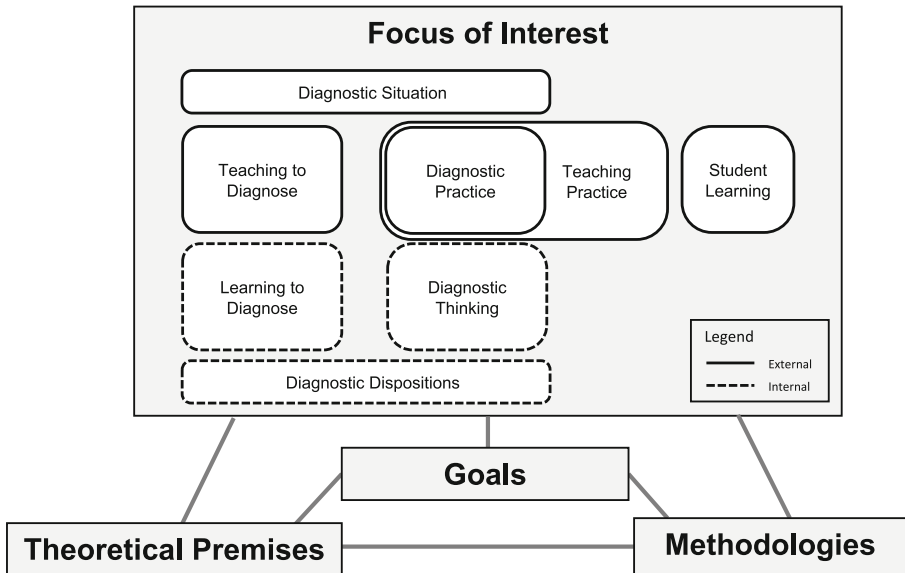


Fig. 1 Overarching framework for research on teachers' diagnostic thinking and practice. The *dashed lines* indicate internal constructs and processes (i.e., within the teacher). The *continuous lines* mark external constructs and behavior

2 An Overarching Framework for Diagnostic Thinking and Practice

Coming from different projects with different research perspectives (NeDiKo, Cosima, and DiaCoM), our aim was to find structuring elements that help to locate heterogeneous perspectives within a common framework. Against this background, we propose an overarching framework for research on teachers' diagnostic thinking and practice that allows for a broader systematic view by structuring research through explicating the respective (1) foci of interest, (2) goals, (3) methodologies, and (4) theoretical premises (Fig. 1). In the following, we use the term *research perspective* to refer to the complex entity formed by these four elements.

To arrive at this framework we, first, extracted various foci of interest from existing research. We chose the foci as a starting point because our research, as that of others, predominantly departs from content related questions. After all authors had agreed on the extracted foci of interest and summarizing descriptions for each of them, we discursively structured them in a *research map* (upper box in Fig. 1). In a second step, we analyzed the literature for additional discriminating and systematizing elements of research on diagnostic thinking and practice that go beyond specific foci of interest. Based on research on diagnostic thinking and practice as well as research-theoretical literature (Price et al. 2017; Grossman and McDonald 2008), we arrived at the following additional elements: goals, methodologies, and theoretical premises. These elements were too diverse to be summarized meaningfully within the visualization of the overarching framework. Therefore, the visualization

only includes placeholders for these aspects (see the middle and lower boxes in Fig. 1).

2.1 Foci of Interest

Current research on teachers' diagnostic thinking and practice has a wide range of foci of interest that can be located on a research map (Fig. 1, upper box). Studies have investigated:

- diagnostic thinking (e.g., Loibl et al. 2020; Krolak-Schwerdt et al. 2013),
- diagnostic practice (e.g., Furtak et al. 2018; Heitzmann et al. 2019),
- the accuracy of diagnostic judgments and its dependence on situations and teachers' dispositions (e.g., Herppich et al. 2010; Südkamp et al. 2012),
- (self-)descriptions of teachers' diagnostic knowledge, skills, beliefs, and other dispositions (e.g., DeLuca et al. 2018; Mertler 2004),
- the impact of diagnostic skills or practices on student learning (e.g., Lee et al. 2020; Kingston and Nash 2011),
- prospective teachers' learning to diagnose (e.g., Codreanu et al. 2020, 2021),
- and effective mechanisms to teach future teachers to diagnose (Heitzmann et al. 2019; Sommerhoff et al. *in press*).

The core of the map consists of the closely connected pair diagnostic thinking and diagnostic practice. Both are accompanied by multiple other central foci of interests. They are divided into foci that are internal to the teacher (dashed lines in Fig. 1) and foci that are external, i.e. visible to others (continuous lines in Fig. 1).

The *external* side contains the diagnostic situation, diagnostic practice, teaching to diagnose, and student learning. *Diagnostic practice* refers to the behavior of the teacher (Heitzmann et al. 2019; Südkamp et al. 2012), such as the gathering of information on the learning conditions, the learning process, or the learning outcome, for example by formal testing, describing observations, or evaluating students' writings. The goal of diagnostic practice is to gain valid knowledge about individual students, small groups, or the whole class—for example, to plan further individual support or whole-class teaching or to inform students and parents.

The *diagnostic situation* sets the frame for possible diagnostic practice and may refer, for example, to the overall context (e.g., a classroom situation or a face-to-face diagnostic interview), the purpose of the diagnosis itself, and instruments that can be employed (e.g., the availability of certain learning materials to demonstrate the learners' current skills or the availability of specific diagnostic tests). The diagnostic situation also includes the specific (mathematical) contents and learning objects that students are supposed to (have) learn(ed) and which are subject to the diagnosis. Moreover, it includes information (e.g., a student's characteristics, behavior, or solution to a specific task) that the teacher uses to reach a judgment. In this sense, the diagnostic situation is regarded as including all aspects relevant to describe, interpret, or explain a students' specific behavior when working on a specific task. The diagnostic situation comprises the students' (hypothetical or actual) knowledge and thinking and the (anticipated or actual) behavior in certain situations (tasks). The description of specific diagnostic situations allows for asking (and possibly answering)

questions such as: What are the relevant misconceptions in a mathematical content area? What are the typical errors and which assessment tasks can elicit them? Leuders and Loibl (2021) argue that these learner- and content-related specificities of the diagnostic situation constitute the content specificity of diagnostic judgments.

In contrast, diagnostic dispositions, diagnostic thinking, and learning to diagnose as *internal* elements of the research map refer to the processes, traits, or states (cognitive or affective) of teachers that underlie diagnostic practice. These elements are psychological constructs that cannot be directly observed and therefore only indirectly empirically supported by measurement or experimental variation. *Diagnostic thinking* refers to the cognitive processes that determine the genesis of judgments, such as perceiving and interpreting information and subsequent decision making (Loibl et al. 2020). These judgments may become visible in teachers' diagnostic practice. Although diagnostic thinking and practice are closely connected, it depends on the research perspective whether one considers practice as regulated by thinking (Schoenfeld 2010) or thinking and practice as inextricably connected processes (as referred to by Furtak et al. 2018). *Diagnostic dispositions* include traits such as teachers' knowledge, beliefs, motivation, and emotions (cf. DeLuca et al. 2018; Looney et al. 2017; Xu and Brown 2016). In addition, they include affective states, such as deliberative or implemental mindsets. In most research, these dispositions are seen as core prerequisites for diagnosis, as they provide, for example, the necessary knowledge for corresponding diagnostic processes. Sufficient content knowledge (CK; see Förtsch et al. 2018; Shulman 1986) as well as pedagogical content knowledge (PCK)—both matching the (mathematical) contents focused in a specific diagnostic situation—are often seen as prerequisite to identify students' (mathematical) errors, misconceptions, learning difficulties, or problematic beliefs. Diagnostic dispositions may not only influence diagnostic thinking and practice but also how diagnostic skills are acquired while *learning to diagnose*.

Generally, teachers' diagnostic practice aims at identifying and possibly optimizing *student learning*. Student learning can only be achieved by a close connection between diagnostic practice and *teaching practice* because the data gathered from diagnosing can be understood as a requirement for effective teaching. This connection adds the complexity of the whole classroom to the focus of research (Seidel and Shavelson 2007; Binder et al. 2018).

2.2 Goals

Strongly interrelated with the different research foci are the (research) goals—that is, the objectives researchers aim to achieve through their research. Although in principle manifold, research in the social sciences can generally be classified based on the goals of *describing*, *predicting*, or *explaining* thinking and behavior (Price et al. 2017). In the field of diagnostic thinking and practice, these goals can be specified at different grades of granularity. For example:

- understanding and predicting teachers' diagnostic judgments;
- measuring interindividual differences in dispositions and skills pertaining to diagnostic practice;

- testing the applicability of general theories to judgments and decision making in pedagogical situations to explain the genesis of judgments;
- developing, evaluating, and optimizing settings to foster diagnostic judgments.

2.3 Methodologies

Within research on teachers' diagnostic thinking and practice, one can find a broad spectrum of methodologies (e.g., case studies, design research, randomized controlled trials, and experiments, using qualitative, quantitative, or mixed methods). The methodologies are associated with the research foci and goals in different ways; some methodologies are better suited to deal with certain foci and to attain specific goals than others. For example, explorative methods are better suited for research with the goal of describing and explaining but less so for predicting. However, there is no simple one-to-one relation between goals and methodologies (e.g., a better understanding of the genesis of teachers' diagnostic judgments can be reached by testing theories experimentally or reconstructing diagnostic thinking).

2.4 Theoretical Premises

The selection of the focus of interest and methodology is explicitly or implicitly based on the theoretical premises that guide the research. For example, experimental research on cognitive processes assumes that teacher judgments can be explained by the interplay of situational and personal characteristics and by the establishment of general theories of human thinking and behavior, ignoring in part the complexity of many (classroom) teaching situations. Similarly, classroom research is mostly based on the theoretical premise that teaching practices can be regarded as constitutive for an explanation of teaching quality and student learning. As theoretical premises can be manifold and context-specific, there is, at least so far, no systematization. However, a decisive distinction is whether these premises are made explicit (implying their consideration by the researchers) or whether they are implicit and possibly opaque for researchers and future research.

2.5 Preliminary Nature and Limitations of the Framework

Having introduced the overarching framework and its key elements, we would like to stress that the presented framework is not intended to provide a comprehensive and systematic review of the *findings* in the field of teachers' diagnostic thinking and practice, but rather a *structural view* on the field. Furthermore, the current version of the framework may need further refinement and elaboration. For example, although we provide an overview of the pivotal aspects with respect to possible foci of interest, we do not suggest a classification for the other elements of the overarching framework (i.e., goals, methodology, and theoretical premises), as we will—in the following analysis—use them first and foremost for describing specific studies in more detail and for providing hints to explicitly reflect on these issues.

3 Research Perspectives—An Analysis of Recent Approaches

The overarching framework aims to provide a strategic orientation for research and to provide impulses for the development of and reflection on research on teachers' diagnostic thinking and practice. The various research perspectives can be specified and delimited from each other by the elements of the overarching framework (i.e., foci of interest, goals, methodologies, and theoretical premises). Thus, for any specific research project, the overarching framework raises the question of how this project specifies and relates elements of the framework to each other and to what extent additional (i.e., so far insufficiently considered) elements of the framework need to be considered theoretically or empirically to further advance research. If, for example, researchers focus exclusively on investigating teachers' diagnostic practices, they make explicit or implicit assumptions about the underlying diagnostic thinking, the embedding of diagnostic practice in the teaching practice, or the diagnostic situation (including the diagnostic goal). We argue that research on diagnostic thinking and practice would benefit from being aware of these assumptions, as validating these assumptions will be an important future step in research.

In the following section, by discussing a selection of different research perspectives, we exemplify how the overarching framework can help to characterize these perspectives (aim 1) and identify directions for future research (aim 2).

To select the examples, we applied the following procedure: First, we analyzed several research perspectives and pertaining studies for their primary foci in a discursive way to arrive at the overarching framework and, in particular, at the research map (cf. Sects. 2 and 2.1). We checked these perspectives for their goals, methodologies, and theoretical premises. Second, we chose four perspectives with different primary foci. Each perspective was coded by one of the authors for its foci and analyzed regarding their goals, methodologies, and theoretical premises. The codings were double-checked by at least one other author and any discrepancies were resolved by discussion. The examples presented in this section illustrate a variety of different research foci and partly contrasting or complementary goals, methodologies, and theoretical premises. In the analyses of these examples, we elaborate on the implicit premises underlying different research perspectives and their interactions with foci of interest and methodologies.

3.1 The “Diagnostic Competence” Research Perspective: Explaining Effective Diagnostic Practice

Within the European context, the term *competence* has been established as one of the core constructs in research on learning (Hartig et al. 2008) and serves as a *research focus* in the literature on teachers' diagnosing of students. The main *goals* within this research perspective are the description, explanation, prediction, and support of teachers' abilities in diagnosing students. Furthermore, the impact of high diagnostic competence on student learning has been investigated (Herppich et al. 2018; Schrader 2011).

The research area is highly heterogeneous, as it is based on different research traditions (Blömeke et al. 2015; Klieme et al. 2008). High competences are often

conceptualized as being successful in dealing with a set of situations (Ufer and Leutner 2017). It differs, however, to what extent researchers focus on either the dispositions necessary for being competent or on the actual behavior of competent people (for a prominent attempt to bring both approaches together, see e.g. Blömeke et al. 2015). Accordingly, some research on *diagnostic competence* focuses more on teachers' diagnostic thinking (e.g., Krolak-Schwerdt et al. 2013), while other studies focus more on teachers' diagnostic practice (e.g., the meta-analysis by Südkamp et al. 2012). In a current overview of the field, Herppich et al. (2018) suggested—in congruence with Blömeke et al. (2015)—explicitly modelling teachers' cognitive dispositions, judgment processes, and performance in situations focusing on diagnosing student characteristics (e.g., Herppich et al. 2018) or tasks (e.g., McElvany et al. 2009). Of particular interest in research on diagnostic competence are informal diagnoses in teaching contexts (Schrader and Praetorius 2018).

In many empirical studies, the *methodological approach* focuses exclusively on measuring teachers' judgment accuracy as an indicator of diagnostic competence (e.g., the meta-analyses by Südkamp et al. 2012; Machts et al. 2016; Urhahne and Wijnia 2021). In these studies, teacher judgments with respect to usually one student characteristic (mostly achievement in mathematics or language, but also other characteristics such as self-concept) are assessed. Teachers' accuracy is then measured based on a comparison to a test, questionnaire, or expert rating. The focus on accuracy has been criticized over the last few years, arguing that the broad concept of competence is too narrowly assessed with a focus entirely on judgment accuracy (e.g., Helmke 2010; van Ophuysen 2010) and questioning the relevance of accuracy for teaching behavior in general (e.g., Abs 2007). Accordingly, other methodologies have been used over the last years, among others, experimental approaches (e.g., Krolak-Schwerdt et al. 2013) or qualitative studies based on case scenarios (e.g., Klug et al. 2013).

The core *theoretical premises* of this research perspective have been made explicit in the definition of teachers' "assessment competence"—which is the international equivalent of the term "diagnostic competence," common in German-speaking countries—by Herppich et al. (2018): "measurable cognitive disposition that is acquired by dealing with assessment demands in relevant educational situations and that enables teachers to master these demands quantifiably in a range of similar situations in a relatively stable and relatively consistent way" (p. 185). Thus, it is assumed that competence (a) can be measured, (b) is mainly cognitive (although some researchers also include other aspects, such as motivation), and (c) needs to be consistently visible in different situations, that (d) these situations need to be pivotal to the educational context (i.e., real-life tasks), that (e) diagnosing is not a goal in itself but is tailored to informing educational decisions and actions, and that (f) there needs to be a certain time stability in competence performances, although diagnostic competence can also be acquired. Other premises that come into play in *measuring* diagnostic competence (for an overview, see Praetorius et al. 2017) are the assumptions that the task selected for measuring diagnostic competence leads to teacher behavior that is somehow representative of teachers' judgments in real-life teaching, that diagnosing student characteristics and teaching these students can be successfully distinguished, and that the chosen approaches for measuring a complex

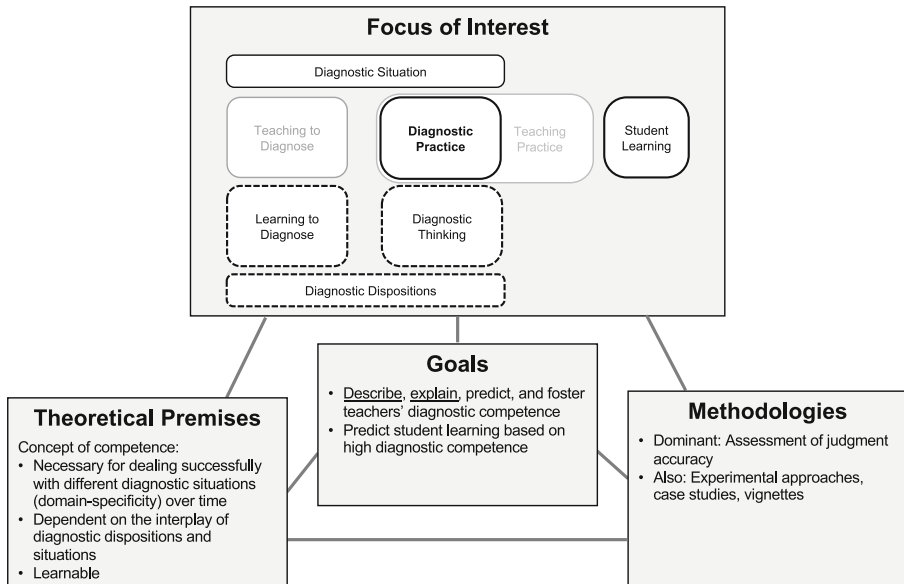


Fig. 2 The “diagnostic competence” research perspective in the overarching framework

construct such as diagnostic competence (e.g., using accuracy) allow their proper display. The foci of interest, goals, methodologies, and theoretical premises are summarized in Fig. 2.

An example from this area of research is a paper by Helmke and Schrader (1987), which, together with the dissertation monograph by Schrader (1989), represented the starting point of research on diagnostic competence in German-speaking countries. Schrader (1989) typically used the term “diagnostic competence” in the German monograph, whereas Helmke and Schrader (1987) replaced it with “judgment accuracy” in the international publication, although both works present exactly the same data. Helmke and Schrader (1987) argued that effective teaching requires an adaptation to the student’s characteristics; hence, diagnosing these characteristics is a prerequisite for adaptation. They further claimed that high-quality diagnoses and high-quality teaching are required for promoting student achievement and that both interact to achieve this positive effect. To test their assumptions, they used data from 690 fifth graders and 32 teachers. Teachers were asked to predict their students’ achievement in a concurrently administered standardized achievement test in mathematics, which focused on the content covered during the last three months of their mathematics instruction. Teachers’ judgments were assessed by asking them to state the number of problems in the mathematics test that they expected each student in their class to solve. The teachers could look at the items of the test before providing their judgments. Judgment accuracy was calculated as a class-wise correlation between the students’ test results and the teacher’s judgments. Teaching was measured using a low-inference observation tool, which was applied to five sequences of five minutes of a lesson over the course of nine lessons. In particular, the authors focused on teachers’ structuring cues (i.e., attention-regulating remarks emphasizing

relevant information and cues with respect to working methods for certain tasks) and individualized supportive contact (i.e., the amount of teachers' individualized contact with students). Judgment accuracy on its own did not make a difference for student achievement. However, such an effect was shown for the combination of judgment accuracy and the two teaching variables.

With respect to the overarching framework, one can conclude that the study focused on teachers' diagnostic practice (in terms of their judgment accuracy), teachers' teaching practice (i.e., structuring cues and individualized supportive contact), and student learning (i.e., a standardized, curricular-valid mathematics test). Teaching and learning how to diagnose did not play a role in this publication. Diagnostic thinking, as well as situations and dispositions, played a crucial role but were not the focus of the topic under investigation. The following premises underlie this study: estimating the mere number of tasks solved correctly by each student is (a) relevant to classroom teaching and (b) representative and sufficient to characterize teachers' diagnostic competence; (c) situational and contextual variables do not influence teachers' diagnoses to a considerable degree; and (d) judgment accuracy is an appropriate measure for teachers' diagnostic competence.

A specific strength of the diagnostic competence perspective is its potential to explain why teachers are diagnosing the way they do. At least theoretically, the competence construct integrates teachers' dispositions, diagnostic thinking, and diagnostic practice with respect to pivotal teaching and learning situations and allows for the explication of their relations. At the same time, research in this area focuses strongly on explaining the differences between teachers, theoretically and methodologically. This narrow focus can be considered a weakness. The perspective may profit from considering how diagnostic competence is acquired (e.g., Klug et al. 2018; see also perspective 3.4). In addition, although research on diagnostic competence focuses on teachers' informal diagnoses, accuracy measurements often necessitate measuring teachers' diagnoses as approximations of formal tests. Thus, the repertoire of methods used in research on diagnostic competence might be inspired by the tools used in cognitive modeling research and research on teachers' diagnostic practice (see perspectives 3.2 and 3.3).

3.2 The “Cognitive Modeling” Research Perspective: Exploring and Explaining the Genesis of Diagnostic Judgments

Scholars in education have called for research on teachers' diagnostic thinking and practice, which explores the thinking underlying teachers' judgments (Herppich et al. 2018; Schrader 2011; Spinath 2005; Leuders et al. 2017). Research with a *focus* on the genesis of diagnostic judgments strives to theoretically explain and empirically investigate the cognitive processes underlying diagnostic judgments based on characteristics of the situation and the teacher (Becker et al. 2020; Rieu et al. 2022). More precisely, these studies focus on what kind of information of the diagnostic situation is perceived by a teacher (e.g., characteristics of the student or features of the task) and how this information is processed (more or less based on knowledge) to derive a judgment that becomes visible in teachers' diagnostic practices (e.g., giving a grade or selecting a task). Regarding the research map, this perspective

focuses on the element of diagnostic thinking and uses the relationships between diagnostic thinking and diagnostic practice, diagnostic dispositions, and diagnostic situations as instrumental for its methodology in the following way: Cognitive models on diagnostic thinking predict how characteristics of the diagnostic situation and dispositions of the teacher result in specific diagnostic practices.

The principal research *goal* within this perspective is to reveal explanatory knowledge on the genesis of diagnostic judgments (Leuders et al. 2020; Loibl et al. 2020) and provide hints on how to design instruction that fosters diagnostic competence and reduces judgment biases (see perspective 3.4 for research focusing explicitly on pre-service teacher education in this field).

As diagnostic thinking is not observable, the following *methodologies* are common in this area: theoretical assumptions on cognitive processes may be generated by exploring external indicators for cognitive processes, such as think-aloud protocols or eye movement (Becker et al. 2020; Philipp 2018), or by detecting correlational patterns between situation or person characteristics and observed behavior (Binder et al. 2018; McElvany et al. 2009). Based on such an analysis and/or by applying existing theories on human reasoning (summarized in, e.g., Bless and Greifeneder 2017) to a diagnostic situation, potential explanations for certain phenomena in diagnostic judgment or, more generally, models for cognitive processes underlying diagnostic judgments can be proposed. These models can then be verified empirically through the experimental manipulation of situations and/or dispositions and testing of whether teachers show the predicted behavior (e.g., Becker et al. 2020;

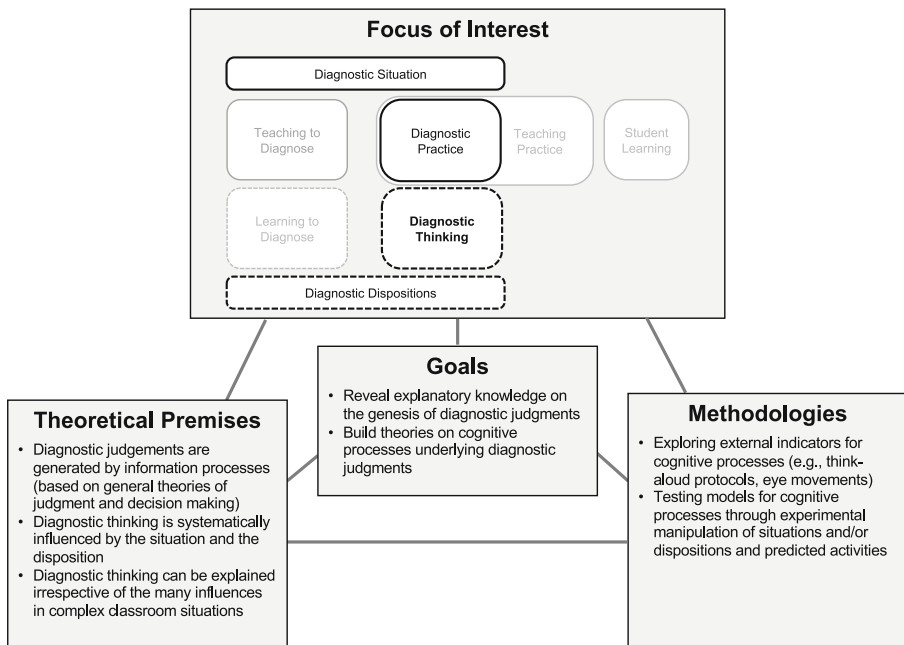


Fig. 3 The “cognitive modeling” research perspective in the overarching framework

Rieu et al. 2022; for a framework underlying this research perspective, see Loibl et al. 2020).

This perspective is based on the *theoretical premise* that diagnostic judgments are generated by information processes (Leuders and Loibl 2021; Loibl et al. 2020), which can be modeled by recurring to the general processes of judgment and decision-making (e.g., knowledge-based processes, top-down/bottom-up processes; see Bless and Greifeneder 2017). Further premises are that these processes are influenced by the situation and teachers' dispositions, which allow manipulation of the processes experimentally, and that diagnostic thinking can be explained irrespective of the complex classroom situation with its many influences. In consequence, any assumptions of the researchers regarding the possible impact of diagnostic practices on teaching or student learning remain implicit. The focus of interest, goals, methodologies, and theoretical premises are summarized in Fig. 3.

To exemplify this research perspective, we provide a short summary of a study by Rieu et al. (2022). The authors focused on diagnostic thinking processes in the context of task diagnosis—that is, when assessing the difficulty of given tasks.

To form a diagnostic judgment, the teacher must identify the relevant information in the diagnostic situation (here, the task features) and integrate this external information with their knowledge (i.e., part of their diagnostic dispositions). Depending on the complexity of the diagnostic situation (e.g., the number of relevant task features), the judgment process may require more or less time. To model and investigate the assumed judgment processes, the study experimentally examines the influence of (1) PCK (by a short intervention on relevant task features and their effect on task difficulty) and (2) the availability of sufficient time on the accuracy of judging task difficulty (as an indicator of the result of the cognitive processes). In line with the hypotheses, the results show that specific PCK is the basis for identifying *and* integrating task features, whereas time pressure affects only the complex process of integrating multiple task features. Regarding the overarching framework, this study investigates diagnostic thinking (perceiving, identifying, and integrating relevant features) by manipulating the diagnostic situation (available time) and the diagnostic disposition (PCK) and experimentally testing assumptions on diagnostic thinking and its results (resulting judgments regarding task difficulty).

This study exemplifies how research from this perspective builds on general theories from social cognition about the formation of judgments and decisions to model the processing of information in the context of diagnostic judgments. A fundamental premise of this type of research is that assumptions on internal, unobservable thinking processes can be tested by means of experimental manipulations of influencing factors. However, the assumptions on learning to diagnose, which underlie the design of the intervention, remain implicit, as learning to diagnose is not the focus of this research (see perspective 3.4 for a complementary approach).

This research perspective has the potential to reveal explanatory knowledge on the genesis of diagnostic judgments. However, in this context, controlled experimental designs require simple operationalizations of the situation, dispositions, and practices. The implicit premise is that, due to the rather generic models on information processing, the findings can be transferred to different situations, contexts, and tasks without further testing. In addition, when extending the view on adjacent

elements (learning to diagnose or teaching practice), it remains speculative to what extent these findings can be transferred to a more realistic context of learning or enacting diagnosing. Furthermore, research from this perspective usually neglects the relationship between diagnostic judgments and pedagogical decisions (teaching practice). Nevertheless, associated pedagogical decisions may also influence judgments.

From this research perspective, interventions are designed to manipulate teachers' knowledge to back up the cognitive model and not to produce long-term effects on teachers' diagnostic competences. Nevertheless, these interventions may still provide indications for formulating principles for promoting diagnostic skills. By taking both perspectives (cognitive modeling and teacher training) into account, future research could investigate how the explanatory knowledge generated by experimentally tested cognitive modeling can be informative for the design of teacher training.

During the last decade, the study of diagnostic thinking has also become a focus of research on diagnostic competence (e.g., Krolak-Schwerdt et al. 2013). This overlap seems to be a promising starting point for linking research that uses general cognitive theories to explain diagnostic thinking with research that departs from complex competence constructs to explain diagnostic practice and effective diagnosing in general (perspective 3.1).

3.3 The “Formative Assessment” Research Perspective: Integrating Diagnostic Practice in Teaching

Under the umbrella term “formative assessment,” an extensive body of research *focuses* on teachers' diagnostic practice and related teaching practice that foster students' learning. Diagnostic practice can range

- from informal thought-eliciting questions (e.g., Ruiz-Primo and Furtak 2007)
- via individual tasks and materials that help to make student understanding visible (e.g., quizzes; Wiliam et al. 2004) or help students to self-diagnose their knowledge (e.g., rubrics; Andrade et al. 2010)
- to systems of such tasks and materials, which are formally embedded into the curriculum (Shavelson et al. 2008).

Associated teaching practices may be any practice that departs from where the students currently are (regarding understanding, motivation, etc.) and promotes learning, often toward a previously defined learning goal (Black and Wiliam 2009). Diagnostic practices are seen as prerequisites for teaching practices and student learning. In related studies, teachers (and students) are taught diagnostic practices. Student learning is often studied. Aspects of the diagnostic situation, such as the discipline or the age of the students, are sometimes studied as moderators of the effects and effectiveness of formative assessment (Kingston and Nash 2011; Lee et al. 2020). Teachers' diagnostic dispositions, particularly their knowledge, are sometimes regarded as prerequisites for good diagnosis and teaching (PCK, Burkhardt and Schoenfeld 2019) or as inextricably interwoven with practices (PCK-in-use, Furtak et al. 2018), but usually not measured. Diagnostic thinking has not been studied (Fig. 4).

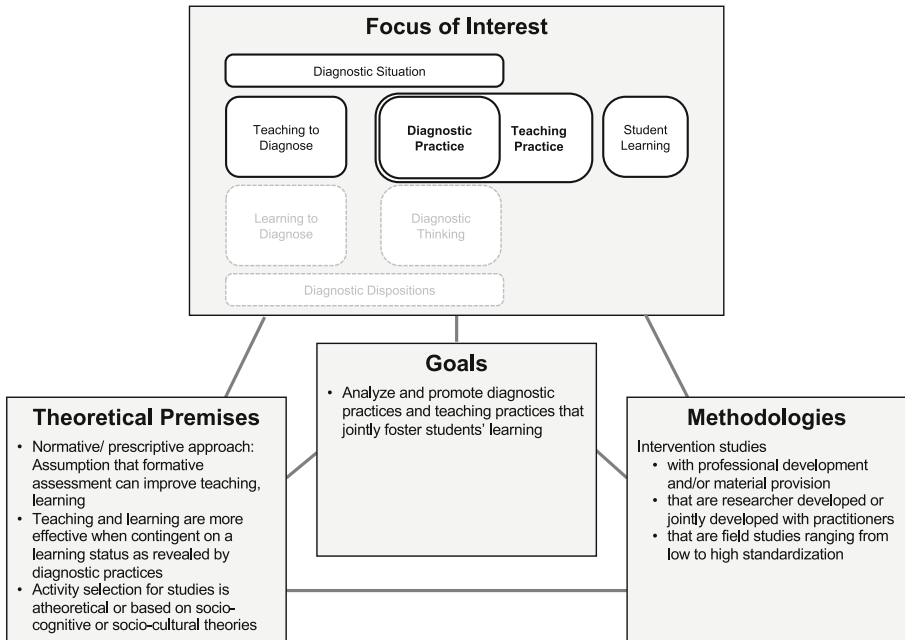


Fig. 4 The “formative assessment” research perspective in the overarching framework

Accordingly, the *goal* of research within this formative assessment perspective is to describe diagnostic practices (e.g., Black and Wiliam 1998) and explain how they relate to and presumably improve student learning (e.g., Lee et al. 2020) in the context of teaching and learning (for overviews, see also, e.g., Black and Wiliam 2009; Kingston and Nash 2011).

Methodologies usually comprise intervention studies, where an intervention is a means to implement diagnostic and teaching practices, but is not necessarily analyzed itself or contrasted with alternative interventions (Lee et al. 2020; Wiliam 2019). Interventions often include professional development (Wiliam et al. 2004) or the provision of teaching materials (Burkhardt and Schoenfeld 2019) or both (Furtak et al. 2018; Kingston and Nash 2011). Interventions may be mainly researcher-developed (Rakoczy et al. 2017) or developed jointly with teachers in a design-based research approach (Wiliam et al. 2004). There is also research on formative assessment that focuses only on the student without involving the teacher (Lee et al. 2020) although such studies are outside the scope of this article. Research projects usually aim at implementation in the field, as opposed to the laboratory, at least as a final goal (Burkhardt and Schoenfeld 2019; Rakoczy et al. 2017), and study designs range from low (e.g., Wiliam et al. 2004) to high standardization (e.g., Rakoczy et al. 2017). Accordingly, studies track various objectives, such as exploration of the field, hypothesis generation, and hypothesis testing, with a focus on testing hypotheses on the effectiveness of formative assessment. The paradigm can be qualitative (Kim 2017), quantitative (Shavelson et al. 2008), or often mixed-methods (Furtak et al. 2018; Wiliam et al. 2004).

The central *theoretical premise* explicit to this research perspective is that teaching practices are more effective in promoting learning when they are contingent on the current learning status as revealed by diagnostic practices (Black and Wiliam 2009; Ruiz-Primo and Furtak 2007). Accordingly, the perspective is normative or prescriptive, as it asks what teaching and learning, including diagnostic practices, *should* be like (Wiliam 2019) and tries to foster these favorable practices. The selection of practices for interventions in early studies has been pragmatic and based on eclectic empirical findings (Wiliam et al. 2004). This research has been joined by more recent projects that rely on coherent socio-cognitive or socio-cultural approaches to teaching and learning (Shepard et al. 2017). These works depart from, but build on, disciplinary theories of learning—for instance, assumptions about misconceptions and conceptual change in mathematics and natural science—to develop practices (Alonzo 2018; Burkhardt and Schoenfeld 2019; for a critique, see Wiliam 2019). They also draw on the views of students as the owners of their learning and the teacher as an engineer of and a coach in the learning situation (Burkhardt and Schoenfeld 2019; Panadero et al. 2018). Another theoretical premise implicit in this perspective is that formative assessment can be learned by teachers and students and can be successfully integrated into teaching and learning, given a benevolent context. The foci of interest, goals, methodologies, and theoretical premises are summarized in Fig. 4.

The description of a classical study (King's-Medway-Oxfordshire Formative Assessment Project [KMOFAP], e.g., Wiliam et al. 2004) may serve to exemplify this research perspective: The researchers designed and improved diagnostic and teaching practices collaboratively with 24 mathematics and science teachers in a series of professional development meetings over 1.5 years. The teachers in the study were provided with knowledge about the general principles of formative assessment and suggested evidence-based practices, “such as rich questioning ..., sharing criteria with learners, and student peer-assessment” (Wiliam et al. 2004, p. 54). The teachers included these principles and practices in their plans on which practices to integrate in their teaching and how to integrate them. The plans were then used to guide teaching practices with a class at the beginning of a new school year. The researchers qualitatively analyzed the plans to examine how teachers translated general recommendations to teaching and observed lessons to examine how teaching gradually changed. In addition, they matched the study classes with comparison classes and found that the intervention had a small-to-medium effect on students' learning outcomes measured by summative assessments (school assessments and a high-stakes test).

In light of the overarching framework, this study reports on the nature of diagnostic practices (e.g., teacher questioning, pre-tests) integrated with teaching practices (e.g., comment-only marking, group work) and their effect on student learning. Examinations took place in a variety of diagnostic situations (teachers from different schools with different classes) but with mathematics and science being similar disciplines as a common feature and without putting a research focus on the diagnostic situations. To enable the analyses, an intervention was implemented; that is, the teachers were “taught” to diagnose. This study exemplifies how research from this perspective focuses on the externally perceivable elements of the framework.

This focus and the related methodology could be a consequence of the normative/prescriptive goal of improving teaching and learning with the help of formative assessment. Against this background, it may be more obvious to concentrate on practices in particular situations than on dispositions and thinking, and this focus might be easier to communicate to the teachers and students involved. In addition, this approach has been proven to be relatively effective (Lee et al. 2020).

Accordingly, an affordance of this research perspective is certainly its potential to inform high-quality teaching and learning. This perspective provides evidence of the diagnostic practices, as well as teaching and learning practices, that are most beneficial for learning and clarifies the conditions under which they are effective. This prescriptive approach, however, might also narrow down the view on activities that are worth studying (Wiliam 2019). This is because there is less interest in describing the activities that are already practiced in teaching and learning situations.

Moreover, from an analytical point of view, this research perspective sheds light on the complex relationships between diagnostic and teaching practices and student learning (Lee et al. 2020). Although several studies analyze these practices within a specifically designed instructional unit (Burkhardt and Schoenfeld 2019; Wiliam et al. 2004), it is sometimes difficult to separate the effects of diagnostic practices integrated with contingent teaching practices from the effects of non-contingent but generally well-designed teaching (Wiliam 2019).

Attending to the foci of interest of the research perspective, views on formative assessment could nevertheless be enriched if researchers considered teacher-internal constructs and processes. For example, an analysis of the knowledge (Shulman 1986) that underlies effective formative assessment as a disposition could help disentangle the degree to which formative assessment is domain-general versus domain-specific (Andrade et al. 2019). Similarly, analyzing teachers' learning to diagnose may be conducive to finding optimal approaches to professional development for formative assessment (perspective 3.4). Moreover, considering teachers' dispositions in combination with their diagnostic thinking may be a way to explain how different teachers use formative assessment and why some are more successful than others (Kim 2017; Wiliam et al. 2004; see perspectives 3.1 and 3.2). In addition, research on formative assessment could profit from a closer description of the applied interventions to optimize diagnostic (and teaching and learning) practices (focus of interest "teaching to diagnose"). These endeavors can yield information about variables that moderate the effectiveness of formative assessment: for example, discipline or type of intervention (Lee et al. 2020; Wiliam 2019; see also perspective 3.4).

3.4 The "Teacher Training" Research Perspective: Supporting Pre-Service Teachers' Diagnostic Thinking and Practice

Research on supporting diagnostic thinking and practice has so far mostly focused on in-service teachers (e.g., Busch et al. 2015; Moyer and Milewicz 2002), while corresponding research on pre-service teachers is scarce. However, findings suggest that pre-service teachers, as well as novel in-service teachers, often have severe difficulties in assessing and diagnosing their students (Codreanu et al. 2021; Levin et al. 2009; Morris 2006). These results resonate with the repeated call for a practice

shift in pre-service teacher education and the creation of so-called “approximations of practice” (Grossman et al. 2009; Heitzmann et al. 2019). To answer these calls and shortcomings, there is currently an upsurge of research on how pre-service teachers’ diagnostic thinking and practice can be supported—for example based on the notion of teachers’ professional vision (Codreanu et al. 2021; Stürmer et al. 2013a), diagnostic activities (Heitzmann et al. 2019), the PID framework (perception, interpretation, decision-making; Blömeke et al. 2015), or research on formative assessment (see Sect. 3.3). Corresponding research mostly *focuses* on either (i) generating empirical and theoretical insights into the teaching and learning of diagnostic skills in the context of university education (e.g., Kron et al. *in press*; Santagata et al. 2007; Stürmer et al. 2013a) or (ii) the development and evaluation of effective instructional formats that support pre-service teachers in acquiring diagnostic skills (Prilop et al. 2021; Sommerhoff et al. *in press*; Stürmer et al. 2013b). With regard to the overarching research framework, this research perspective thus centrally focuses on the elements of teaching to diagnose (e.g., different support mechanisms and learning environments; Prilop et al. 2021; Yeh and Santagata 2015), learning to diagnose (e.g., as a theoretical premise, a mediator for learning outcomes based on certain measurable process characteristics, or a non-observable cognitive process), and diagnostic practice as the central outcome variable (Codreanu et al. 2021; Santagata et al. 2007; Stürmer et al. 2013b). Pre-service teachers’ diagnostic dispositions are often considered either as a control variable or as a means to explain the differential effectiveness of support mechanisms (Sommerhoff et al. *in press*; Stürmer et al. 2013a). Accounting for the diagnostic situation helps to generalize findings or compare effects between situations—for example, between different subjects, such as biology and mathematics—or even with diagnostic skills in other professions, such as medicine (Chernikova et al. 2020a). Diagnostic thinking as a non-observable cognitive process is mostly only implicitly considered or approximated via the pre-service teachers’ displayed diagnostic practice. As studies are usually concerned with pre-service teachers, explicit teaching practice or student learning are generally not addressed further, and outcome measures mostly relate to pre-service teachers’ diagnostic practice or its output (e.g., patterns within diagnostic practice, written diagnosis, or judgment accuracy).

The *goal* of the corresponding research is to better describe and explain the acquisition and development of diagnostic skills, especially focusing on the effectiveness of different types of support and conditions for their effectiveness throughout teacher education.

The *methodologies* employed within this research perspective can be either (i) effect-focused or (ii) design-focused. Effect-focused methods comprise mostly experimental and quasi-experimental empirical intervention studies (e.g., Besser et al. 2015; Chernikova et al. 2020a, b), which allow for an at least partially controlled setting and robust, generalizable empirical evidence. However, corresponding approaches often do not focus on the creation, refinement, and later use of instructional materials. In contrast, design-focused methods (e.g., design-based research), focus specifically on the creation and refinement of instructional materials, often based on multiple iterative testing cycles. Generated data is then used to advance research on supporting pre-service teachers’ diagnostic thinking and practice; however, re-

sults are often less generic and bound to specific situations such as, for example, the instructional materials themselves. As of today, only a few research projects have reached the point where they have combined (or even tried to combine) both methodologies (e.g., Glogger et al. 2013).

The central *theoretical premise* underlying this research perspective is that pre-service teachers’ diagnostic practice is positively related to in-service teachers’ diagnostic practice and positively impacts student learning (see also premises in the context of formative feedback). Other premises are mostly related to (i) the differential effectiveness of certain support mechanisms (see, e.g., Belland et al. 2017; Chernikova et al. 2020b), (ii) more or less general learning theories (e.g., cognitive load theory), and (iii) the effects of support mechanisms on diagnostic thinking and practice. These premises underlie some studies within this research perspective either implicitly or explicitly, or they are explicitly investigated in others (mostly premises (i) and (iii)). The foci of interest, goals, methodologies, and theoretical premises are summarized in Fig. 5.

The teacher training perspective and its classification based on the overarching framework can be exemplified based on the study by Heitzmann et al. (2018). They examined the effects of error-explanation prompts and adaptable feedback as two types of scaffolds on different forms of diagnostic knowledge, forming the basis of diagnostic competence. Based on results from prior research regarding the use of worked examples (Renkl and Atkinson 2010), the authors point to the advantages and possible disadvantages of error-explanation prompts by building on cognitive load theory and the concept of knowledge encapsulation (Kalyuga 2011). Moreover, the authors based their design decisions on prior results regarding

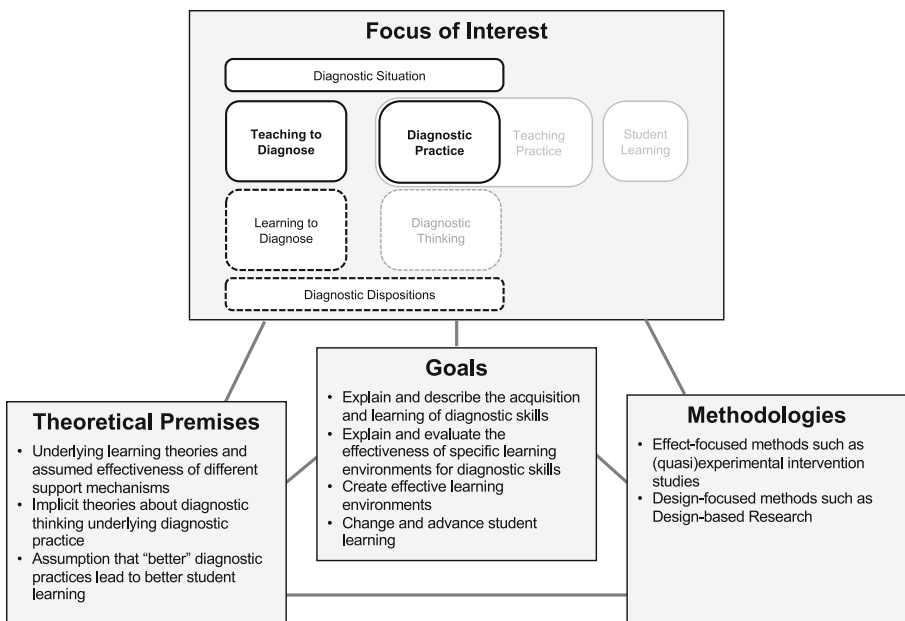


Fig. 5 The “teacher training” research perspective in the overarching framework

adaptable feedback and self-explanation prompts, which have shown positive effects on diagnostic competences in medical education (Stark et al. 2011). Using a 2×2 factorial design, Heitzmann et al. (2018) examined the effects of error-explanation prompts and adaptable feedback and found significant interactions of both types of support and nuanced impacts of either type of support on the different outcome variables. In contrast to their hypotheses, neither type of scaffold for learning to diagnose showed a clear positive effect, possibly due to the additional demand the self-explanation prompts posed.

With regard to the overarching framework, this study focuses on diagnostic practice (used as process data and as outcome variable), participants' diagnostic learning (inferences about participants' learning based on pre-test/post-test effects), and teaching to diagnose (modes of action of the examined scaffolds and general conclusions about scaffolds) while controlling for the diagnostic situation (standardized across research conditions) and participants' dispositions (prior knowledge). The goal of this research is clearly effect-focused, trying to shed light on the differential effectiveness of certain scaffolds using an experimental design as methodology. The explicit theoretical premises, some of which could not be verified empirically, were related mainly to the effectiveness of the scaffolds, while the implicit premises concerned the transformation of diagnostic knowledge into diagnostic practice in authentic learning and teaching situations. The results of the study highlight, in particular against the research map for the research foci, that insufficient attention was given to diagnostic thinking and diagnostic practice (in the sense of process measures) a priori, yet they needed to be considered at least partially to interpret the results a posteriori.

One potential benefit of this research perspective is that it not only provides empirical evidence on learning and teaching diagnostic skills—that is, on how to effectively support pre-service teachers in improving their diagnostic thinking and practice—but also yields corresponding artifacts, such as video-based simulations (Fischer and Opitz 2022). Often, these simulations can be implemented more or less directly to innovate university teaching and are thus highly valuable for practice. To this end, a combination of effect- and design-focused methodologies would be instrumental. However, based on the missing link between teacher practice and student learning, even research results uncovering highly effective ways of supporting pre-service teachers remain partially specious, as it is mostly an assumption that these learning gains transfer to later teaching situations and student learning. In particular, the magnitude of the effects on student learning is currently unclear; educational interventions in pre-service teacher training may lead to only negligible effects in later teaching situations. As there is currently no evidence in this regard, longitudinal studies also involving pre-service teachers' later practice are necessary.

Moreover, while theoretical assumptions regarding learning to diagnose and the non-observable diagnostic thinking in the context of this research perspective are often highly plausible, these assumptions are often either based on (i) theoretical considerations, (ii) findings from areas other than diagnostic thinking and practice, or (iii) little evidence. Therefore, it currently remains an open question whether and how this research perspective should aim at a more systematic consideration of learning to

diagnose and diagnostic thinking as two further aspects of the overarching research framework, or if it should be informed by other perspectives (e.g., perspective 3.2).

4 Discussion

Research on teachers' diagnostic thinking and practice is highly diverse and includes various perspectives that have evolved from different research traditions, disciplines, and cultural backgrounds. Today, there is only little overview of the differences and commonalities of these research perspectives, particularly because they often make use of different wording and constructs, which hampers research syntheses.

To advance research on teaching and teacher education, Grossman and McDonald (2008) emphasized the need for developing a common language and structure that allows bringing together different research perspectives and favors fruitful communication and knowledge exchange. To address this call in the context of teachers' diagnostic thinking and practice, the present paper presents an overarching framework that includes suggestions with respect to core terminology in the field. These suggestions are based on extended discussions among the authors. Moreover, the proposed overarching framework provides a structure for the field to stimulate more critical reflection on our own studies. The four structuring elements of the overarching framework are (a) foci of interest, (b) goals, (c) methodologies, and (d) theoretical premises. The first is core to the framework, as it provides a systematic overview of key foci of interest in research on teachers' diagnostic thinking and practice. The other three allow for describing research with different foci of interest in more depth and make explicit the guiding forces in research. Although there might be additional elements that could be important, we believe these four elements already allow for an initial systematic overview of the field and single empirical studies.

Applying the overarching framework to example studies from four core perspectives in research on teachers' diagnostic thinking and practice, we tested its usefulness in identifying the strengths and limitations of previous research and its potential to inform future research. We were able to categorize all four examples regarding the overarching framework, and these categorizations differed to a considerable degree, providing evidence that the framework is useful in distinguishing different research perspectives. Our analyses also revealed the strength of the overarching framework in identifying aspects that have not been sufficiently considered (or at least not sufficiently communicated) in single studies and within different research perspectives so far. Combining the examples, all aspects distinguished with respect to possible foci of interest could be identified—some of which were covered in several perspectives (e.g., teachers' diagnostic practice), whereas others only in one of the perspectives (e.g., diagnostic thinking). However, no single study covered all of them and, more importantly, explicitly addressed the decisions with respect to the chosen focus/foci. The fact that pivotal aspects often remain implicit is even more obvious for the theoretical premises of all research perspectives addressed in this paper. These premises were neither sufficiently considered in presenting the research desiderata nor in interpreting the results of the studies. In some works, they are used for post-hoc interpretations of results. We are convinced that, to advance re-

search, we all need to be more rigorous in presenting the assumptions we make right from the beginning and also present evidence on why we believe these assumptions are justified. This would raise awareness of open issues and implicit assumptions in our own work and help make our research more cumulative, empirically, and theoretically.

We already commented on the preliminary nature of the overarching framework with its strong focus on research foci. The current version appeared to be sufficiently detailed and systematic for an analysis of the exemplary studies in Chap. 3. However, to achieve a systematic overview of the field, the analyses need to be further substantiated by conducting a systematic review of the entire field. Only then can one fully judge the potential of the overarching framework.

We did not conduct such a systematic review so far, as we deem it necessary to first agree on a common language and structure (Grossman and McDonald 2008) so that the review will help advance the field to the largest extent possible. We, therefore, invite other researchers to advance our suggestion in such ways that it helps our community most. As a first step in this direction, the editors of the current special issue locate all papers in this issue within the overarching framework in their introduction to the special issue. Doing so allows for a more extended test of the capability of the framework to unfold the similarities and differences among these papers and identify current desiderata in research on teachers' diagnostic thinking and practice. In addition, we invite all researchers to contribute to this discussion (e.g., in papers and symposia) with the goal of refining our collective understanding of the research field.

Acknowledgements Part of the research was funded by the Ministry of Science, Research, and the Arts Baden-Württemberg, Germany.

Funding Open Access funding enabled and organized by Projekt DEAL.

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

References

- Abs, H.J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern [Reflections on modeling diagnostic competence in teachers]. In M. Lüders & J. Wissinger (Eds.), *Forschung zur Lehrerbildung* (pp. 63–84). Münster: Waxmann
- Alonzo, A.C. (2018). Exploring the learning progression–formative assessment hypothesis. *Applied Measurement in Education*, 31(2), 101–103
- Andrade, H.L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17(2), 199–214
- Andrade, H.L., Bennett, R.E., & Cizek, G.J. (Eds.). (2019). *Handbook of formative assessment in the disciplines*. Routledge

- Becker, S., Spinath, B., Ditzgen, B., & Dörfler, T. (2020). Der Einfluss von Stress auf Prozesse beim diagnostischen Urteilen – eine Eye Tracking-Studie mit mathematischen Textaufgaben [The influence of stress on processes in diagnostic judgments—an eye tracking study with mathematical text tasks]. *Unterrichtswissenschaft*, 48, 531–550
- 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
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Besser, M., Leiss, D., & Klieme, E. (2015). Wirkung von Lehrerfortbildungen auf Expertise von Lehrkräften zu formativem Assessment im kompetenzorientierten Mathematikunterricht [Effect of teacher training on teachers' expertise on formative assessment in competency-based mathematics education]. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*, 47(2), 110–122
- Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic skills of mathematics teachers in the COACTIV study. In *Diagnostic competence of mathematics teachers* (pp. 33–53). Cham: Springer
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31
- Bless, H., & Greifeneder, R. (2017). General framework of social cognitive processing. In R. Greifeneder, H. Bless & K. Fiedler (Eds.), *Social cognition: how individuals construct social reality* (pp. 16–36). London, New York: Routledge
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3–13
- Burkhardt, H., & Schoenfeld, A. (2019). Formative assessment in mathematics. In H. L. Andrade, R. E. Bennett & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 35–67). Routledge
- Busch, J., Barzel, B., & Leuders, T. (2015). Promoting secondary teachers' diagnostic competence with respect to functions: development of a scalable unit in Continuous Professional Development. *ZDM*, 47(1), 53–64
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020a). Facilitating diagnostic competences in higher education—a meta-analysis in medical and teacher education. *Educational Psychology Review*, 32, 157–196
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020b). Simulation-based learning in higher education: a meta-analysis. *Review of Educational Research*, 20(10), 1–43
- Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2020). Between authenticity and cognitive demand: finding a balance in designing a video-based simulation in the context of mathematics teacher education. *Teaching and Teacher Education*, 95, 103146
- Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2021). Exploring the process of preservice teachers' diagnostic activities in a video-based simulation. *Frontiers in Education*, 6(133), 626666. <https://doi.org/10.3389/educ.2021.626666>
- DeLuca, C., Lapointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25, 355–375
- Fischer, F., & Opitz, A. (2022). *Learning to diagnose with simulations—examples from teacher education and medical education*. New York: Springer
- Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. (2018). Systematizing professional knowledge of medical doctors and teachers: development of an interdisciplinary framework in the context of diagnostic competences. *Education Sciences*, 8(4), 207. <https://doi.org/10.3390/educsci8040207>
- Furtak, E. M., Bakeman, R., & Buell, J. Y. (2018). Developing knowledge-in-action with a learning progression: sequential analysis of teachers' questions and responses to student ideas. *Teaching and Teacher Education*, 76, 267–282

- Glogger, I., Holzäpfel, L., Kappich, J., Schwonke, R., Nückles, M., & Renkl, A. (2013). Development and evaluation of a computer-based learning environment for teachers: assessment of learning strategies in learning journals. *Education Research International*, 2013(1), 1–12
- Grossman, P., & McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *American Educational Research Journal*, 45(1), 184–205
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P.W. (2009). Teaching practice: a cross-professional perspective. *Teachers College Record*, 111(9), 2055–2100
- Hartig, J., Klieme, E., & Leutner, D. (Eds.). (2008). *Assessment of competencies in educational contexts*. Hogrefe
- Heitzmann, N., Fischer, F., & Fischer, M.R. (2018). Worked examples with errors: when self-explanation prompts hinder learning of teachers' diagnostic competences on problem-based learning. *Instructional Science*, 46, 245–271
- Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M.R., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., & Fischer, F. (2019). Facilitating diagnostic competences in simulations in higher education: a framework and a research agenda. *Frontline Learning Research*, 7(4), 1–24
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts [Teaching quality and teachers' professionalism. Diagnosis, evaluation and improvement of teaching]*. Seelze: Klett
- Helmke, A., & Schrader, F.W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98
- Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2010). Do tutors' content knowledge and beliefs about learning influence their assessment of tutees' understanding? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 314–319). New York, NY: Erlbaum
- Herppich, S., Praetorius, K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193
- Kalyuga, S. (2011). Cognitive load theory: how many types of load does it really need? *Educational Psychology Review*, 23(1), 1–19
- Kim, H. (2017). Teacher learning opportunities provided by implementing formative assessment lessons: becoming responsive to student mathematical thinking. *International Journal of Science and Mathematics Education*, 17, 341
- Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. *Assessment of Competencies in Educational Contexts*, 3, 22
- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: a process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, 30, 38–46
- Klug, J., Schultes, M.T., & Spiel, C. (2018). Assessment at school—Teachers' diary-supported implementation of a training program. *Teaching and Teacher Education*, 76, 298–308
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: a social cognitive analysis. *Social Psychology of Education*, 16, 215–239
- Kron, S., Sommerhoff, D., Ahtner, M., Stürmer, K., Wecker, C., Siebeck, M., & Ufer, S. (2022). Cognitive and motivational dispositions behind diagnostic competences of pre-service teachers: combined effects of professional knowledge and interest on the diagnostic process and accuracy. *Journal für Didaktik der Mathematik*. <https://doi.org/10.1007/s13138-022-00200-2>
- Lee, H., Chung, H.Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: a systematic review. *Applied Measurement in Education*, 33(2), 124–140
- Leuders, T., Loibl, K. (Eds.) (2021). Beyond subject specificity—student and teacher thinking as sources of specificity in teacher diagnostic judgments [Special Issue]. *RISTAL*, 4, 60–70
- Leuders, T., Dörfler, T., Leuders, J., & Philipp, K. (2017). Diagnostic competence of mathematics teachers: unpacking a complex construct. In T. Leuders, T. Dörfler, J. Leuders & K. Philipp (Eds.), *Diagnostic competence of mathematics teachers. Unpacking a complex construct in teacher education and teacher practice* (pp. 3–32). New York: Springer

- Leuders, T., Loibl, K., & Dörfler, T. (2020). Diagnostische Urteile von Lehrkräften erklären – Ein Rahmenmodell für kognitive Modellierungen und deren experimentelle Prüfung [Explaining diagnostic judgments of teachers—a framework for cognitive models and their experimental testing]. *Unterrichtswissenschaft*, 48, 493–502
- Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education*, 60(2), 142–154
- Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic Judgements by cognitive modeling (Diacom). *Teaching and Teacher Education*, 91, 1–10. <https://doi.org/10.1016/j.tate.2020.103059>
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25, 442–467
- Machts, N., Kaiser, J., Schmidt, F. T., & Moeller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: a meta-analysis. *Educational Research Review*, 19, 85–103
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Ullrich, M., et al. (2009). Diagnostische Fähigkeiten von Lehrkräften [Diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie*, 23(34), 223–235
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education*, 33, 49–64
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749
- Morris, A. K. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Mathematics Teacher Education*, 9(5), 471–505
- Moyer, P. S., & Milewicz, E. (2002). Learning to question: categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education*, 5(4), 293–315
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz – eine theoretische und empirische Annäherung. [Professional pedagogical-diagnostic competence—a theoretical and empirical approach]. In N. Berkemeyer, W. Bos, H. G. Holtappels, N. McElvany & R. Schulz-Zander (Eds.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Vol. 16, pp. 203–234). Juventa
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *Australian Journal of Educational Research*, 45, 13–31
- Philipp, K. (2018). Diagnostic competences of mathematics teachers with a view to processes and knowledge resources. In T. Leuders, T. Dörfler, J. Leuders & K. Philipp (Eds.), *Diagnostic competence of mathematics teachers. Unpacking a complex construct in teacher education and teacher practice* (pp. 109–127). Cham: Springer
- Praetorius, A.-K., Hetmanek, A., Herppich, S., & Ufer, S. (2017). Herausforderungen bei der empirischen Erforschung diagnostischer Kompetenz [Challenges in empirical research of diagnostic competence]. In A. Südkamp & A.-K. Praetorius (Eds.), *Diagnostische Kompetenz von Lehrkräften – Theoretische und methodische Weiterentwicklungen [Diagnostic competence of teachers—theoretical and methodological developments]* (pp. 95–150). Münster: Waxmann
- Price, P. C., Jhangiani, R. S., Chiang, I. A., Leighton, D. C., & Cuttler, C. (2017). *Research methods in psychology* (3rd edn.). Pullman, WA
- Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2021). The role of expert feedback in the development of pre-service teachers' professional vision of classroom management in an online blended learning environment. *Teaching and Teacher Education*, 99, 103276
- Rakoczy, K., Klieme, E., Leiss, D., & Blum, W. (2017). Formative assessment in mathematics instruction: theoretical considerations and empirical results of the Co2CA project. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Eds.), *Competence assessment in education: research, models, and instruments* (pp. 447–467). Berlin: Springer
- Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. L. Plass, R. Moreno & R. Brünken (Eds.), *Cognitive load theory* (pp. 91–108). New York: Cambridge University Press
- Rieu, A., Leuders, T., & Loibl, K. (2022). Teachers' diagnostic judgments on tasks as information processing—the role of pedagogical content knowledge for task diagnosis. *Teaching and Teacher Education*, 111, 103621

- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84
- Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: an empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, 10(2), 123–140
- Schoenfeld, A. H. (2010). *How we think: a theory of goal-oriented decision making and its educational applications*. New York: Routledge
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts [Diagnostic competencies of teachers and their importance for the design and effectiveness of teaching]*. Frankfurt am Main: Peter Lang
- Schrader, F.-W. (2011). Lehrer als Diagnostiker [Teacher as diagnosticians]. In E. Terhart, H. Bennewitz & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 683–698). Münster: Waxmann
- Schrader, F.-W., & Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern [Diagnostic competence of parents and teachers]. In D. H. Rost, J. R. Sparfeldt & S. Buch (Eds.), *Beltz Psychologie 2018. Handwörterbuch pädagogische Psychologie* (5th edn., pp. 92–98). Weinheim, Basel: Beltz
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21, 295–314
- Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, 98(6), 47–52
- Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14
- Sommerhoff, D., Nickl, M., Codreanu, E., Ufer, S., & Seidl, T. (in press). Pre-service teachers' learning of diagnostic skills in a video-based simulation: Effects of conceptual vs. interconnecting prompts on judgement accuracy and the diagnostic process. *Learning and Instruction*
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teachers' assessment of student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95
- 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
- Stürmer, K., Könings, K. D., & Seidel, T. (2013a). Declarative knowledge and professional vision in teacher education: effect of courses in teaching and learning. *British Journal of Educational Psychology*, 83(3), 467–483
- Stürmer, K., Seidel, T., & Schäfer, S. (2013b). Changes in professional vision in the context of practice. *Gruppendynamik und Organisationsberatung*, 44(3), 339–355
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: a meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762
- Ufer, S., & Leutner, D. (2017). Kompetenzen als Dispositionen – Begriffsklärungen und Herausforderungen [Competencies as dispositions—explanations of terms and challenges]. In A. Südkamp & A.-K. Praetorius (Eds.), *Diagnostische Kompetenz von Lehrkräften – Theoretische und methodische Weiterentwicklungen* (pp. 67–74). Münster: Waxmann
- Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32, 100374
- Van den Heuvel-Panhuizen, M. H. A. M. (1996). *Assessment and realistic mathematics education*. Vol. 19. Utrecht: Utrecht University
- Wiliam, D. (2019). Conclusion: why formative assessment is always both domain-general and domain-specific and what matters is the balance between the two. In H. L. Andrade, R. E. Bennett & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 243–264). Routledge
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, 11(1), 49–65
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teaching and Teacher Education*, 56, 149–162
- Yeh, C., & Santagata, R. (2015). Preservice teachers' learning to generate evidence-based hypotheses about the impact of mathematics teaching on learning. *Journal of Teacher Education*, 66(1), 21–34