A Learning Analytics Approach to Address Heterogeneity in the Classroom: The Teachers' Diagnostic Support System

Tobias Kärner¹ · Julia Warwas² · Stephan Schumann¹

Published online: 19 May 2020
© The Author(s) 2020

Abstract

Addressing heterogeneity in the classroom by adapting instruction to learners' needs challenges teachers in their daily work. To provide adaptive instruction in the most flexible way, teachers face the problem of assessing students' individual characteristics (learning prerequisites and learning needs) and situational states (learning experiences and learning progress) along with the characteristics of the learning environment. To support teachers in gathering and processing such multidimensional diagnostic information in class, we have developed a client-server based software prototype running on mobile devices: the Teachers' Diagnostic Support System. Following the generic educational design research process, we (1) delineate theoretical implications for system requirements drawn from a literature review, (2) describe the systems' design and technical development and (3) report the results of a usability study. We broaden our theoretical understanding of heterogeneity within school classes and establish a basis for technological interventions to improve diagnostic accuracy in adaptive instructional strategies.

Keywords Student heterogeneity · Diagnostic support · Situational assessment · Learning analytics · Educational technology

1 Introduction

Given a student body heterogeneous in both the cognitive and emotional-motivational prerequisites of learning, teachers face the challenge of providing adaptive instruction. While the educational literature reveals a broad consensus on this general statement and discusses various instructional strategies to create adaptation and differentiation (e.g. Hertel et al. 2011), comparably less attention is devoted to which factors make teachers' instructional decisions particularly challenging, how teachers usually arrive at their decisions, which dimensions of "heterogeneity" they should take into account and how they can be

✉ Tobias Kärner
tobias.kaerner@uni-konstanz.de

¹ Business and Economics Education, Department of Economics, University of Konstanz, Konstanz, Germany

² Business and Economics Education, Faculty of Business and Economics, University of Goettingen, Goettingen, Germany

technically supported in achieving well-founded instructional decisions. We argue that the quality of teachers' decisions about adaptive instructional measures greatly depends on the accessibility and validity of multidimensional and even time-varying diagnostic information. During ongoing instructional processes, however, teachers' mostly informal diagnostic activities are vulnerable to judgement biases because they have to integrate information from various sources and evaluate spontaneous impressions of behavioural cues to adjust instructional strategies flexibly under time pressure. The task of continually collecting, integrating and displaying valid diagnostic information can be executed by a technology-based Teachers' Diagnostic Support System (TDSS). The system aims to enhance the accuracy of teachers' judgements about individual learning potentials and needs while students deal with the current learning environment. TDSS can thus lay an important foundation to justify teachers' decisions on "adapting" instructional strategies for a given group of learners in a dynamic learning environment.

When describing the steps of the development process, we rely on the *generic educational design research process* that is common in the field of educational technology development (McKenney and Reeves 2014, 2019). According to this approach, "educational design research is a genre of research in which the iterative development of solutions to practical and complex educational problems provides the setting for scientific inquiry" (McKenney and Reeves 2014, p. 131). The research process therefore evolves in "iterated cycles of design, enactment, and analysis" (Sandoval 2013, p. 389). According to McKenney and Reeves (2014, 2019), three phases can be distinguished within the process of educational design research: first, an analysis/orientation phase; second, a design/development phase; and third, an evaluation/retrospective phase. The analysis/orientation phase includes problem definition, as well as exploration and review of existing theoretical approaches, previous technical developments and empirical findings via literature reviews, context analysis and needs assessment. The design/development phase includes exploration and mapping of solutions, as well as building and revising solutions within the processes of system construction and technical implementation. The evaluation/retrospective phase includes planning data collection, collecting and analysing data, as well as reflection about the evaluation results. Because of the cyclic and iterative character of the design process, these three main phases are often revisited during the lifespan of a project. The evaluation results can thus be used as starting points for further progress.

In the second section, we provide a detailed theoretical analysis and literature review (*phase 1 in the educational design research process*). To substantiate the system's utility and functions, we first elaborate the generic dimensions that constitute heterogeneity in the classroom, namely interindividual variability across students within a given class, intraindividual variability within students and interactions between students and features of the learning environment. Additionally, we review evidence on teachers' diagnostic abilities to gain an understanding of how well teachers usually manage to form accurate judgements of these dimensions. We then attend to issues of situational complexity that complicate the formation of accurate diagnostic judgements in ongoing instructional processes. Furthermore, we highlight explanatory models of how teachers process diagnostic information and situational assessment to theoretically underpin the development of our support system. The third section of the paper outlines the general design features of the prototype TDSS and the technical aspects of system development (*phase 2 in the educational design research process*). In this section, we draw conclusions from the preceding literature review for system requirements, define system functions and workflows and describe the technical implementation and user interface design. In the fourth section, we report the findings of a usability study (*phase 3 in the educational design research process*). The user-oriented

evaluation of the TDSS deployed an adapted version of IsoMetricsS (Gediga et al. 1999; Hamburg and Gediga 2002), and we recruited 14 practising teachers to run representative tasks to uncover potential problems or errors with software usage and data representation. The paper concludes with an outlook on the practical implications and implications for further research.

2 Theoretical Analysis and Literature Review

2.1 Heterogeneity in the Classroom and the Limited Accuracy of Teachers' Diagnostic Judgements

Implementing adaptive instruction necessitates highly differentiated and valid knowledge about the specific factors that constitute the heterogeneous learning potentials and needs in a given class. Teachers have to assess students' learning processes, as well as the individual and situational characteristics that may affect these processes (Brühwiler and Blatchford 2011). The diagnostic tasks that teachers must fulfil to accomplish this are not restricted to formalized and occasional assessments of learning prerequisites or outcomes at a given point in time. To *adapt instructional strategies* to learners' needs, teachers need to monitor students' learning experiences, classroom communication behaviours and other factors that contribute to or indicate learning progress continually. Because routine sampling of student data is a common method in educational research (e.g. Sembill et al. 2008), it has also been implemented in formative assessment approaches that make frequently measured student data available to teachers (e.g. Black and Wiliam 1998; Dunn and Mulvenon 2009; Kingston and Nash 2011; Spector et al. 2016; Van der Kleij et al. 2015). In a broader sense, the educational use of continually sampled data is located in the field of *learning analytics*, which "uses dynamic information about learners and learning environments, assessing, eliciting and analyzing it, for real-time modeling, prediction and optimization of learning processes, learning environments, as well as educational decision-making" (Ifenthaler and Widanapathirana 2014, p. 222).

We have identified three main sources of heterogeneity in the literature that should inform teachers' judgements and decisions on adaptive instruction: (1) *interindividual* variability (such as different cognitive abilities) among students within a class (e.g., Geisler-Brenstein et al. 1996), (2) *intraindividual* variability within students (such as time-varying progression rates in dealing with instructional contents and building conceptual understanding; e.g. Faber et al. 2017), and (3) *interactions* between student characteristics and instructional features (such as interactions of different levels of student knowledge with different degrees of task difficulty; e.g. Corno and Snow (1986), see the summary in Kärner et al. (2019a, b), too).

Research on teachers' diagnostic competencies also suggests that most teachers lack the ability to assess sources and degrees of heterogeneity among learners correctly. Most related studies focus on formalized and selective evaluations of students' learning preconditions or proficiency levels (at the beginning and end of a unit or school year) or relatively stable personal characteristics. Thus, extant research mostly concerns judgement accuracy regarding *interindividual variability*. In sum, the available evidence suggests that this accuracy ranges from small to moderate effect sizes, depending on the object of diagnosis. Recent meta-analytic findings (e.g. Südkamp et al. 2012) indicate that teachers are rather skilled in assessing students' domain-specific proficiency levels. Comparably lower

accuracy is reached when gauging students' general cognitive abilities (Machts et al. 2016). Markedly weaker congruence exists between students' actual emotional, motivational and psychosocial characteristics and teachers' assessments of these characteristics. Assessment objects include, for instance, school-related anxiety (Spinath 2005; Südkamp et al. 2017), well-being (Urhahne and Zhu 2015), interest (Hosenfeld et al. 2002), motivation (Praetorius et al. 2017; Spinath 2005; Urhahne et al. 2013), engagement (Kaiser et al. 2013), academic self-concept (Praetorius et al. 2017; Südkamp et al. 2017; Spinath 2005), understanding (Hosenfeld et al. 2002), work habits and social behaviour (Stang and Urhahne 2016) and sympathy-based peer interactions (Harks and Hannover 2017).

Although precise estimations about *intraindividual variability* within students have been deemed an indispensable requisite of teaching quality (cf. van Ophuysen and Lintorf 2013), it has seldom been investigated empirically (e.g. Warwas et al. 2015). Similarly, only a few research contributions (e.g. Karst 2012; Schmitt and Hofmann 2006) have systematically examined teacher judgements about *interactions between students and instructional features* in class. The next section therefore elaborates why continuous monitoring of these time-varying and situation-bound sources of heterogeneity during classwork is a highly demanding task and even one that is prone to more judgement errors than summative evaluations of comparably stable student characteristics.

2.2 Situational Complexity and the Challenges of Situational Judgement during Classwork

As Beck (1996) points out, the concept of an “instructional situation” entails multiple constituents that contribute to potentially divergent interpretations of current professional demands in the view of different teachers: (1) available time for processing diagnostic information (e.g. behavioural cues from students) within a selected (2) physical space (e.g. classroom), including different (3) objects and relationships between objects (e.g. students, subject matter, interactions between students and subject matter) and different (4) social roles (e.g. student, teacher). In addition, (5) pre-existing schemata and scripts of appropriate actions (e.g. pedagogical beliefs) will determine a teacher's perception and comprehension of situational constituents, as well as his/her anticipation of potential courses of development. Furthermore, (6) affective appraisal (e.g. like, dislike) will inhibit or facilitate subsequent actions. Beck (1996) stresses that successful instructional interactions depend on how participants (teachers, students) succeed in coordinating their subjective representations of current situations. In a similar vein, the literature on teachers' diagnostic competencies discusses several difficulties that impede adequate situational judgements during classwork (cf. Warwas et al. 2015):

Information overload. To assess and support students' learning experiences and progress continually, teachers must process information from various, possibly ambiguous, information sources (e.g. students, learning materials and media) simultaneously (Böhmer et al. 2017; Krolak-Schwerdt et al. 2009). Accomplishing this task within a limited time frame places high demands on a teacher's cognitive resources, as human working memory has limited capacity (Cowan 2010). Such information-processing restrictions are discussed in terms of “information overload” (Eppler and Mengis 2004; Schneider 1987) and “cognitive load” (Sweller et al. 2011). In such instances, information reduction that leads to a parsimonious but possibly invalid mental representation of the situation is an eminent aspect of coping with situational complexity and preserving one's capacity for action (cf. Fischer et al. 2012). Particularly under conditions of stress,

which frequently occur in the dynamics of classroom interactions and events, people tend to narrow their field of attention to a small number of (subjectively) salient aspects of the environment, thus adopting a “cognitive tunnel vision” (Endsley 1995). As a consequence, people reach decisions without sufficiently exploiting available information (Endsley 1995). Both attention narrowing and premature conclusions can negatively affect the accuracy of situational judgements and the appropriateness of taken measures (cf. Norman 2009).

Time pressure and schema-driven situational judgements. Although teachers need time to process situational information, not all instructional arrangements allow their (temporary) retreat from ongoing interactions to reflect deliberately about what is going on in detail. While the sheer amount of learning-relevant situational information may sometimes exceed working memory capacity during action execution (see above), time pressures to reach immediate decisions during ongoing interactions with students further impede elaborate planning, well-founded judgements and reasoned strategic choices. This unavoidable “pressure to act” (Weinert and Schrader 1986) forces teachers into highly automated and schematized detections of (1) states, changes and discrepancies of student behaviours, (2) the progress of teaching and learning processes, and (3) the effects of their own actions. In turn, such schema-driven processing of information can lead to overly simplistic and very rigid subjective representations of factually based situational conditions and requirements (cf. Endsley 1995; Stanton et al. 2009).

Validity of situational information. During instruction, teachers normally do not have access to valid and reliable behavioural indicators of psychological student characteristics (e.g. intellectual abilities, academic self-concept) and/or to their learning experiences and progress (e.g. situational understanding of subject matter). Without valid data from standardized tests and students’ explicit self-reports of learning experiences, teachers’ judgements about students’ prerequisites, experiential states and progression in learning are vulnerable to cognitive biases (cf. Förster and Souvignier 2015; Rausch 2013).

Framing effects in evaluations of situational information. The accuracy of teachers’ judgements of student characteristics demonstrably depends on their frame of reference (Hosenfeld et al. 2002; Lorenz and Artelt 2009). Südkamp and Möller (2009), for example, studied the effects of students’ academic achievement on teachers’ assessments. The authors found a negative effect of class average achievement scores on teacher judgements: “students with identical performance were graded more favourable in a class with low average achievement than in a class with higher average achievement” (Südkamp and Möller 2009, p. 161). The relativity of situation-bound information in terms of framing plays a major role in judgement processes (cf. Tversky and Kahneman 1981).

Multidimensionality of situational information. Following an interactionist paradigm, behaviours and experiences are determined by personal as well as context-related characteristics (Kärner et al. 2017; Mischel 1968; Nezlek 2007; Richard et al. 2003). Thus, trying to understand behaviours and experiences only by the main effects of person and/or context variables is too simplistic (Kärner and Kögler 2016; Kärner et al. 2017). For teachers’ diagnostic tasks to assess students’ learning behaviours and experiences, the interactionist paradigm implies that information about individual learners, as well as their current learning environment, must be considered simultaneously and in tandem to decide on appropriate learning support (Schmitt and Hofmann 2006).

Based on the addressed generic determinants of heterogeneity and situational complexity in any classroom, teachers must consider and evaluate a multitude of information on learners and learning environments to reach decisions on adaptive instructional measures. In the following section, we take a closer look at the usual psychological processes, as well

as the desirable analytical steps, for the collection and integration of diagnostically relevant information.

2.3 Acquiring and Processing Information During Class

2.3.1 Situation Assessment and Situation Awareness as a Basis for Continual Situational Judgements

Following a literature review conducted by Dominguez (1994), situation awareness (SA) pertains to the “continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events” (Dominguez 1994, p. 11). Dominguez (1994) further distinguishes the *process* of SA, which “includes extraction, integration, and the use of the mental picture” from the *product* of SA as “the picture itself”. Endsley (1995, p. 36) defines the process of “achieving, acquiring, or maintaining SA” as situation assessment, whereas SA is seen as a state of knowledge. Thus, the quality of SA mainly depends on the quality of situation assessment (Dominguez 1994), which, in turn, depends on the quality of monitoring activities performed to acquire information about the situation (Lau et al. 2012).

There are several theoretical perspectives on SA (Salmon et al. 2008). In the following, we will describe in greater detail two perspectives that seem to be most important for our analyses of SA in teaching and learning contexts. The first perspective stems from the individual-focused theory of Endsley (1995), who pursues an information-processing approach. For the second, we draw on the theory of distributed SA of Stanton and colleagues (2006), who pursue a systemic approach. Endsley (1995) separates three hierarchical phases of achieving and maintaining SA: (1) perception of the attributes, status, and dynamics of relevant environment elements; (2) comprehension based on a synthesis of perceived cues; and (3) projection of a system’s future development. By adapting Endsley’s SA model to teaching activities during class, for instance, the teacher must perceive the behavioural cues of students when they are working on a specific task; if one or more students appear to be clueless, the teacher must change his/her course and give further instructions. In general, the instructor must be capable of perceiving and decoding behavioural cues, linking them to pre-existing concepts and deciding what to do during further teaching and learning.

Whereas Endsley’s (1995) information-processing approach assumes SA to be primarily an individual phenomenon, Stanton and colleagues (2006) assume that SA is distributed across agents within complex collaborative systems and emerges from their interactions (cf. Neisser 1976; Smith and Hancock 1995). The concept of distributed SA assumes, among other things, that different members of a collaborative system have different views on the same situation, and SA holds loosely coupled systems together via communication between members (Stanton et al. 2004, 2006). The distributed SA model also postulates that, within collaborative systems, members have different (but in principle compatible) SA—notwithstanding whether the information to which they have access is the same or different. The way in which members use the information also differs, depending on personalized views on the world; thus, their SAs will not be identical, even when they have access to exactly the same information (Salmon et al. 2008). As Stanton and colleagues (2009) assume the interindividual variability of SA, the authors represent a constructivist view. By adapting the distributed SA model to teaching activities during class, the

classroom can be viewed as a complex collaborative system with teacher(s) and students as agents with different roles, probably different goals and expectations and individual views of the same situation. Salmon and colleagues (2012) state that it is the compatibility between different views that determines the efficiency of the whole system. This, in turn, is in line with Beck's (1996) argument that educational processes mainly depend on how participants (teachers, students) succeed in coordinating their individual representations of current situations. It is also in line with Heid (2001), who states that it is not the "objective situation" that matters but rather the subjective representation of a situation.

2.3.2 Collection and Integration of Diagnostic Information as the Core of Teachers' Diagnostic Activities

Teachers' judgements on individual learning potentials and needs while students deal with the current learning environment are crucial for adapting instructional strategies. In line with Jäger's (2006) diagnostic process model, *learning analytics tools* can support teachers in adopting a hypothesis-driven approach to gathering and interpreting diagnostically relevant information (cf. van Ophuysen and Lintorf 2013). Starting with a specific instructional problem or question (for instance, the impression that some students are mastering exercise units well while others are falling behind), the teacher can deliberately test his/her own assumptions on possible reasons in different learner groups (e.g. deficits in conceptual understanding versus task-specific motivation). Diagnostic activities in a *hypothesis-driven approach* to identifying situationally suitable micro-adaptions of instruction that merge available data on students and context start with a concrete didactical *problem or question* (e.g. Box et al. 2015; Brink and Bartz 2017; Shirley and Irving 2015; Suurtamm et al. 2010). After problem clarification, the teacher must formulate concrete *assumptions* that specify the pursued question and *diagnostic criteria or indicators* to be assessed to test these assumptions (e.g. student knowledge, subjective understanding; Antoniou and James 2014; Cisterna and Gotwals 2018; Hondrich et al. 2016; Ruiz-Primo and Furtak 2007; van den Berg et al. 2016). The teacher also has to define the mode of *variable operationalization* (e.g. standardized tests, rating scales; Förster and Souvignier 2014; Shirley and Irving 2015; van den Berg et al. 2016). Based on his/her decisions about diagnostic indicators and operationalization, the teacher has to plan the procedure for *data collection*, including frequency of assessment, available technical devices and seamless integration of collected data into running teaching and learning processes (e.g. Box et al. 2015; Brink and Bartz 2017). The data collection phase is followed by a phase of *data analysis and interpretation of the findings*. The obtained findings are the basis for *hypothesis testing*, which offers the opportunity to check the plausibility of the effects of the chosen microadaptive strategies in terms of *practical evidence* for their educational value (e.g. motivational or knowledge gains from specific kinds of teacher feedback or specific instructional methods; Box et al. 2015; Cisterna and Gotwals 2018; Hondrich et al. 2016; van den Berg et al. 2016).

Several elaborated *learning analytics frameworks* for collection and integration of diagnostic information exist that incorporate, for instance, different data-flow architectures, data analysis algorithms and modes of data representation (e.g. Greller and Drachsler 2012; Ifenthaler and Widanapathirana 2014; Siemens 2013). Nevertheless, van Ophuysen and Behrmann (2015) stress that there is as yet no integrated technology available that is able to completely support the above-mentioned steps of the individual teacher's diagnostic process and related instructional decisions in the classroom. The review by Kärner and colleagues (2019) of technological support systems for teachers points in a similar direction.

Assessing the functional scope of 66 systems, the study concluded that data- and document-management functions are much more common than functions that promote collaboration and communication or functions that provide support for complex decision-making processes. More precisely, systems that support teachers in making justified diagnostic and microadaptive instructional decisions by providing options for targeted analyses and graphical displays of context- and student-related information are, to our knowledge, currently rare or non-existent.

3 System Design and Technical Development

3.1 Implications from Literature Reviewed for System Requirements

As our outline of the challenges that complicate situational judgements shows, information load and time pressure, as well as the lack of valid or comprehensive situational information, can seriously hamper teachers' continual and mostly informal diagnostic activities during class. Teachers should have the opportunity to gather information about situational characteristics in class whenever needed—that is, based on repeated data samplings during instructional units—to make adequate situation assessments that aid in achieving and maintaining SA (cf. Endsley 1995). As Endsley (1995, p. 35) states, a person's perception of learning-relevant elements in the environment determines his or her SA. Moreover, this awareness will also be a function of the design features of the available support system “in terms of the degree to which the system provides the needed information and the form in which it provides it” (Endsley 1995, p. 35). We can therefore draw some implications for the development of a technology-based support system for teachers' situational diagnostics.

Beck's (1996) dimensions of educational situations provide a basis for the definition of the information that must be collected, processed and provided to the teacher to support him/her in achieving and maintaining SA. First, Beck (1996) assumes time for processing situational information. Therefore, relevant diagnostic information—which, in the case of teachers' continual and informal monitoring of classroom activities, is usually elusive—should be permanently available to allow deliberate and substantiated decisions for adaptive instruction. Physical space and social roles determine the data structure and role management of the support system: a school employs different teachers, who teach in different classes with different students. If we assume that classroom instruction is a complex collaborative system, teachers and students will have different, but in principle compatible, SA (cf. Salmon et al. 2008; Stanton et al. 2006, 2009). If we further assume that the compatibility between the views of different classroom participants determines the efficiency of their collaborative system (in terms of interactive instructional processes), a technology-based support system must provide information about all of the objects involved in these instructional processes, as well as the relationships between these objects (cf. Beck 1996; Salmon et al. 2012). Therefore, information from different sources must be collected and integrated: (1) repeated situation-bound self-reports of students' learning experiences (indicating their subjective perceptions of classroom situations) and repeated tests of students' current knowledge about the instructional content; (2) standardized tests for the assessment of situationally invariant student characteristics and traits; and (3) descriptions of situation-bound instructional characteristics (e.g. media use, teaching and learning goals) rated by the teachers themselves. Multiple data sources allow assessment and monitoring of the interplay of student and instructional characteristics that together affect

students' situational experiential states and steps in their learning progress. Such a triangulation of data from teacher ratings, student self-reports and tests—all referring to specific instructional situations—combined with standardized tests assessing student's relatively stable characteristics, may support teachers in comparing their individual representations of current situations with students' individual situational representations—and thus, to adapt instruction adequately to learners' needs (cf. Beck 1996; Salmon et al. 2012).

3.2 Definition of System Functions and Workflow

Based on the above-mentioned implications for technology-based support of teachers' continual diagnostic activities during class, we developed the Teachers' Diagnostic Support System (TDSS). In reference to Power's (2002) classification of information systems, our software can be labelled a *task-specific decision support system* that provides information relevant for situational assessments and instructional decisions within teachers' specific working contexts (classrooms). Furthermore, our software supports online functions for integrating and analysing multidimensional data from a variety of sources and creating views and data representations via a graphical user interface (e.g. charts and tables; cf. Power 2002).

To support teachers in adapting instruction under consideration of heterogeneity among students, the system functions for the collection and analysis of data are of particular interest. The TDSS allows *data collection* that spans (1) information on students' personal characteristics (e.g. domain-specific knowledge and competencies, emotional-motivational characteristics), (2) descriptions of instructional characteristics (e.g. characteristics of the learning content), and (3) information on students' learning experiences and learning progress (e.g. situational interest in the subject matter, actual knowledge about the topic). *Analytical functions* include analyses of (1) interindividual variability between students within class (e.g. differing prior knowledge), (2) intraindividual variability within students (e.g. states of learning progress at different times), (3) variability of instructional characteristics (e.g. variability of teaching methods across time), and (4) interactions between students and instructional characteristics (e.g. different states of student knowledge and different degrees of task difficulty).

After defining system functions, we can outline the workflow (Fig. 1). Before the teacher uses the system for the first time, the administrator configures the system and inserts initial information about the students' personal characteristics. The teacher then configures the system within his/her permissions. During instruction, the teacher triggers data samplings whenever they are needed immediately or were planned in advance. Information on instructional characteristics (teacher as data source) and on learning experiences and learning progress of students (students as data sources) can be recalled on demand. During and after class, the teacher can retrieve and analyse data.

3.3 Technical Implementation and User Interface Design

The coding of the TDSS software was realized by a professional software development company. For software technical implementation, .NET Core 2.0 was used as the basis for web services, the server was hosted on Microsoft Azure and Angular (compiled to HTML, JS and CSS) was used for front-end implementation. Application programming interfaces were implemented via C#. For database implementation, MSSQL was used, and the charts for graphical data analyses are generated via Chart.js. Technical software development followed

Fig. 1 Phases, workflow and information

an iterative process including prototyping, testing, analysing, revising and refining the software (cf. Nielsen 1993). The TDSS is client–server based software that runs on mobile devices such as smartphones or tablet PCs. The system enables real-time data collection and analysis. Figure 2 illustrates the implemented client–server architecture.

Figure 3 illustrates the process of data abstraction. Information on *students* (learning experiences and learning progress) and *context* (instructional characteristics) are collected during class. In the resulting cross-classified data structure, students' states (learning experiences and learning progress) are nested both within students (characterized by personal characteristics assessed before class) and within context segments (characterized by instructional characteristics assessed during class).

The graphical user interface allows teachers to perform different analyses as defined above. Figure 4 illustrates example views of the user interface design.

Fig. 4 Example views of user interface design

Fig. 4 (continued)

4 Evaluation of System Usability

4.1 Method

According to the third phase of the educational design research process, the development results were evaluated (cf. McKenney and Reeves 2014, 2019). In this phase, the authors and the software coding team first tested the system for logical consistency and functional validity before we asked in-service teachers to evaluate the system's usability.

We considered the following topics within the consistency and validity check: (1) creation and management of users, roles and permissions (administrator, teacher, student), (2) general system configuration (e.g. creating tests), (3) display of graphical user interface and menu navigation, (4) data entry, (5) system functions for real-time data collection and analyses via sample data, (6) real-time request-response checks of client-server communication, and (7) online analytical processing of multidimensional data (e.g. functional consistency and validity of database queries). After software testing, we revised and refined the system functions to minimize errors and to ensure desired system functionality.

For user-oriented software evaluation, we asked representative users to run typical tasks to uncover potential usage problems, ambiguities and errors when dealing with the prototype version of the TDSS (cf. Lazar et al. 2017). The system was equipped with example data for 10 fictional students, who were characterized by individual values for prior knowledge and learning motivation, thus providing information of interindividual variability among students. Information on students' situational interest about the subject matter and actual knowledge about the topic was given with four measures for each student, enabling assessment of intraindividual variability within students. Instructional characteristics were covered via different degrees of student-centred learning activities. Based on the deposited example data, the test persons tested the system functions for data analyses.

4.2 Sample

The sample consisted of 14 practising German teachers (4 males, 10 females) with a mean age of 33.9 years ($SD=6.8$, Min.=28, Max.=53) and a mean work experience of 6.3 years ($SD=4$, Min.=3, Max.=16). Twelve teachers worked at a vocational school, and two teachers worked at a high school. All participants provided written informed consent. A one-hour workshop introduced the background of the study and the software to the teachers. Afterwards, teachers worked on the sample data on their own.

4.3 Measures

For user-oriented software evaluation, we adapted the IsoMetricsS questionnaire covering the design principles of International Organization for Standardization (ISO) 9241 Part 10 (Gediga et al. 1999; Hamborg and Gediga 2002). The following system characteristics were measured on 5-point Likert scales (1 = "predominantly disagree", 5 = "predominantly agree", or "no opinion").¹ In addition, the participants had the opportunity to give further comments and impressions verbally:

¹ Questions on *suitability for individualization* were not assessed for the current version of the TDSS because the system cannot be customized individually at the moment.

- (1) *Suitability for the task* (10 items, $\alpha=0.873$; e.g. “The presentation of information on the screen supports me in performing my diagnostic tasks”)
- (2) *Self-descriptiveness* (6 items, $\alpha=0.622$; e.g. “The terms and concepts used in the software are clear and unambiguous”)
- (3) *Controllability* (7 items, $\alpha=0.818$; e.g. “The software lets me return directly to the main menu from any screen”)
- (4) *Conformity with user expectations* (7 items, $\alpha=0.767$; e.g. “When executing functions, I have the feeling that the results are predictable”)
- (5) *Error tolerance* (8 items, α =not reportable; e.g. “No system errors (e.g. crashes) occur when I work with the software”)
- (6) *Suitability for learning* (6 items, $\alpha=0.888$; e.g. “So far I have not had any problems in learning the rules for communicating with the software, i.e. data query”)

4.4 Findings

4.4.1 ISO Design Principle Ratings

The descriptive findings show that the participants rated all of the assessed design principles as being above average (Table 1). The perceived suitability of the system for conducting continual diagnostic tasks in classrooms was rated quite high ($M=3.73$), and the participants considered the TDSS to be useful for supporting their diagnostic tasks. They also rated positively the self-descriptiveness of the software ($M=4.07$), its controllability ($M=4.23$) and conformity with user expectations ($M=4.50$). The suitability for learning was also assessed positively ($M=4.25$), which is relevant for prospective user trainings. We found the highest ratings for error tolerance ($M=4.55$). However, this scale contained many missing values because the corresponding items refer, for instance, to possible errors with data entry, but the participants only tested the system functions for data analyses.

Table 1 ISO design principle ratings

Assessed design principles	N	Min	Max	M	SD
Suitability for the task	14	3.20	4.22	3.73	0.36
Self-descriptiveness	14	2.80	5.00	4.07	0.59
Controllability	13	3.14	5.00	4.23	0.53
Conformity with user expectations	14	3.00	5.00	4.50	0.54
Error tolerance	4	4.00	5.00	4.55	0.53
Suitability for learning	14	3.00	5.00	4.25	0.69

1 = “predominantly disagree”, 5 = “predominantly agree”

4.4.2 Verbal Statements Concerning System Usefulness and System Optimization

Verbally, the participants described aspects of system usefulness and possible modifications and extensions. In the following, we systematize participants' statements about the challenges of teachers' diagnostic activities during classwork, as set out in the theoretical part of the paper, and the functions for data collection and analysis that were implemented in the TDSS.

Considering classroom instruction as a complex collaborative system, participants considered the TDSS to be useful in reducing the complexity of decisions on differentiated instructional measures for distinct learner groups. Two participants mentioned the potential of system-supported *student grouping* that they considered important, such as for implementing inclusive education. In that regard, a technologically assisted group formation necessitates the definition of attributes that guide the process of composing the groups, specifically member attributes (e.g. dominant learning style or domain-specific knowledge) as well as group attributes (e.g. degree of homogeneity of group member characteristics). In the case of automated assignment of students to groups, the grouping techniques still need to be determined (e.g. fuzzy clustering; Maqtary et al. 2019).

Three participants emphasized the potential of the TDSS to integrate multidimensional situational information, which supports adaptive instruction by allowing individualized assignment of tasks with varying difficulty levels, and by facilitating the detection of possibly differential effects of alternative instructional methods (e.g. partner work vs. individual work) on individual learning outcomes. Both application examples refer to the interaction between student characteristics and learning conditions. In that regard, the TDSS is considered useful in supporting instructional decision-making because it merges *multiple data sources* (student self-reports, standardized tests and descriptions of instructional characteristics).

Besides complexity reduction and data triangulation, some participants noted that it is important that using the TDSS does *not* produce *additional* complexity and information load. Three participants thus wished for improvements in the *ergonomic design* of the teacher user interface (e.g. more effective menu navigation, colour display, wording, information representation, graphs and tabular forms). Two other participants stressed the relevance of ergonomically designing the student user interface as well. Yet another participant suggested creating effective and automated solutions for *importing existing student data* (e.g. grades, class lists) into the TDSS.

Two teachers stressed that both the induction phase and the routine application of a diagnostic support system should be effective and should not take much time. Concerning *time restrictions and time pressure* in instructional processes, it is important to give feasible recommendations for implementing the TDSS within teachers' daily classroom work. We assume that the use of the system would be more effective in student-centred than in teacher-centred phases of instruction because, under the first condition, the teacher is able to retreat from ongoing interactions, has no immediate pressure to act and can reflect about what is going on in detail, based on a thorough inspection of the information provided by the TDSS.

Moreover, the free-text answers reveal that the teachers are well aware of the fact that they often lack access to *valid and reliable student information* that is indispensable for adequately addressing student heterogeneity in classes. Participants mentioned, for instance, the need for information on various performance indicators, preparatory training and educational background, absenteeism rates, social backgrounds, behavioural disorders

and even addiction issues, in order to create learning tasks and materials that meet the differing prerequisites and needs of individual learners or learner groups. Three participants wished to have the option for an *individualized configuration* of the TDSS by generating questions and tests for their students themselves and by integrating such additional items in the data collection process. Four teachers highlighted the usefulness of the TDSS for *plausibility checks* of their subjective student-related or situation-related judgements and subsequent instructional decisions. In this respect, the collected multidimensional data could even enrich *collaborative case discussions in teacher teams*, in addition to existing methods of information triangulation such as 360-degree feedback, thus facilitating multiple perspectives for decision-making. For the same purpose, *information exchange* between teachers, school administration and parents is considered useful for coordinating subjective representations of educational situations (cf. Beck 1996; Salmon et al. 2012). However, with respect to extensive information exchange between different parties outside the single classroom's boundaries, the participants also stressed the importance of compliance with prevailing *data protection regulations*.

5 Summary and Outlook

We introduced a client–server based software prototype to support teachers' daily diagnostic tasks, which can be called a task-specific decision support system within the learning analytics framework. In line with the generic educational design research process (McKenney and Reeves 2014, 2019), we delineated system requirements drawn from a literature review, described the implemented system functions for data collection and analyses and reported the results of a usability study.

To advance the cyclical educational design research process, we aim to take the following steps, the first of which concerns comprehensive field tests of system application. In the reported usability study, the participants only checked the system functions for data analyses, while data collection, consistency and validity checks for data entry were simulated by the software development team and the authors of this paper. Subsequent evaluations must include longer-term usability tests of the TDSS during real-time classroom interactions to establish its practical suitability in teachers' daily work. Second, we used an established self-report instrument (IsoMetricsS) to assess the perceived system usability; however, in subsequent studies, additional evaluation methods such as observation, log file analysis or eye tracking could yield a more nuanced picture of how teachers select and analyse available diagnostic information. In particular, insights gained through these methods could supply the system developers with ideas for improved menu options or graphical displays of complex interactions between student and context characteristics. Third, complementing the teachers' perspective on system usability, the ergonomic design of the TDSS's student interface should also be scrutinized and, if necessary, improved. Fourth, in the more distant future, alternative methods and sources for the collection and integration of diagnostic data could be put to the test. Given sufficiently large numbers of classes, educational data mining algorithms (e.g. for composing learner groups) could be implemented. To reduce interruptions to instructional processes—which inevitably occur whenever students are prompted to give self-reports on their experiential states—biosensitive detectors might also be a noteworthy supplement for the TDSS's database. This would, however, require particularly elaborate feasibility tests that account for both data protection regulations and convergent validity in the assessment of the relevant constructs.

Last but not least, systematic educational effectiveness research is needed to examine the system's capacity to enhance teachers' diagnostic judgements and students' learning activities. With this aim, empirical field studies should be complemented with experimental studies. Referring to previous research on formative assessment (e.g. Black and Wiliam 1998; Dunn and Mulvenon 2009; Kingston and Nash 2011), it seems worthwhile to investigate (a) favourable conditions for system integration into prevalent instructional designs and (b) the effects of system usage on students' learning outcomes. These studies could clarify the instructional decisions and situations for which diagnostic support from the TDSS is most beneficial, if system usage enhances the accuracy of teachers' judgements of students' learning needs, as well as which adaptive strategies teachers derive from given information. Apart from the contextual conditions of system usage, individual teacher competence in reading and interpreting diagnostic data (e.g. Espin et al. 2018)—that is, differences in their *Educational Data Literacy*—must be considered as a potential predictor of the system's educational effectiveness. In particular, longitudinal studies should provide more empirical evidence on the assumed positive relations between a targeted and proficient use of multifaceted diagnostic information, improved microadaptive strategies in instructional processes and student learning progress.

Acknowledgements Open Access funding provided by Projekt DEAL. The software development and the usability study were funded by the Young Scholar Fund of the University of Konstanz (Project-ID: P83949718 FP 497/18 Kärner). The authors would like to thank Jimmy Warwas, Lennart Brunner and Thomas Eidloth-Tiemann from the Warptec Software GmbH (Bamberg, Germany) for the professional coding of the software. Further, the authors would like to thank the anonymous reviewers for their critical, but very helpful and constructive comments and suggestions.

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

- Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability*, 26(2), 153–176.
- Beck, K. (1996). Die «Situation» als Bezugspunkt didaktischer Argumentation – Ein Beitrag zur Begriffspräzisierung [The «situation» as reference point for didactical argumentation – A contribution to clarify the concept]. In W. Seyd & R. Witt (Eds.), *Situation, Handlung, Persönlichkeit. Kategorien wirtschaftspädagogischen Denkens* [Situation, action, personality. Categories of vocational-educational thinking] (pp. 87–98). Hamburg: Feldhaus.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Böhmer, I., Gräsel, C., Krolak-Schwerdt, S., Hörstermann, T., & Glock, S. (2017). Teachers' school tracking decisions. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education. Research, models and instruments* (pp. 131–147). Cham, Switzerland: Springer.
- Box, C., Skoog, J., & Dabbs, G. (2015). A case study of teacher personal practice assessment theories and complexities of implementing formative assessment. *American Educational Research Journal*, 52(5), 956–983.
- Brink, M., & Bartz, D. E. (2017). Effective use of formative assessment by high school teachers. *Practical Assessment, Research and Evaluation*, 22(8), 1–10.

- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108.
- Cisterna, D., & Gotwals, A. W. (2018). Enactment of ongoing formative assessment: challenges and opportunities for professional development and practice. *Journal of Science Teacher Education*, 29(3), 200–222.
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 605–629). New York: Macmillan.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57.
- Dominguez, C. (1994). Can SA be defined? In M. Vidulich, C. Dominguez, E. Vogel, & G. McMillan (Eds.), *Situation awareness: Papers and annotated bibliography* (pp. 5–15). Report AL/CF-TR-1994–0085. Wright-Patterson Air Force Base, Ohio: Air Force Systems Command.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research and Evaluation*, 14(7), 1–11.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Eppler, M. J., & Mengis, J. (2004). The concept of information overload—A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5), 325–344.
- Espin, C. A., Saab, N., Pat-El, R., Boender, P. D. M., & van der Veen, J. (2018). Curriculum-Based Measurement progress data: Effects of graph pattern on ease of interpretation. *Zeitschrift für Erziehungswissenschaft*, 21(4), 767–792.
- Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers and Education*, 106, 83–96.
- Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving*, 4(1), 19–42.
- Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, 32, 91–100.
- Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students' reading progress. *School Psychology Review*, 44(1), 60–75.
- Gediga, G., Hamborg, K.-C., & Düntsch, I. (1999). The IsoMetrics usability inventory. An operationalisation of ISO 9241–10 supporting summative and formative evaluation of software systems. *Behaviour and Information Technology*, 18(3), 151–164.
- Geisler-Brenstein, E., Schmeck, R. R., & Hetherington, J. (1996). An individual difference perspective on student diversity. *Higher Education*, 31(1), 73–96.
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, 15(3), 42–57.
- Hamborg, K.-C., & Gediga, G. (2002). Questionnaire for the evaluation of graphical user interfaces based on ISO 9241/10. <https://www.isometrics.uni-osnabrueck.de/qn.htm>. Accessed 2 October 2018.
- Harks, M., & Hannover, B. (2017). Sympathiebeziehungen unter peers im klassenzimmer: wie gut wissen lehrpersonen bescheid? [Sympathy-based peer interactions in the classroom: how well do teachers know them?]. *Zeitschrift für Erziehungswissenschaft*, 20(3), 425–448.
- Heid, H. (2001). Situation als konstrukt. Zur kritik objektivistischer situationsdefinitionen [Situation as construct. A critical review of objectivist definitions of the situation]. *Schweizerische Zeitschrift für Bildungswissenschaften*, 23(3), 513–528.
- Hertel, S., Warwas, J., & Klieme, E. (2011). Individuelle förderung und adaptive lerngelegenheiten im grundschulunterricht. Einleitung in den thementeil [Individual fostering and adaptive learning opportunities in elementary school instruction. An introduction]. *Zeitschrift für Pädagogik*, 57(6), 854–867.
- Hondrich, A. L., Hertel, S., Adl-Amini, K., & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy and Practice*, 23(3), 353–376.
- Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische kompetenz: unterrichts- und lernrelevante schülermerkmale und deren einschätzung durch lehrkräfte in der unterrichtsstudie SALVE. In M. Prenzel, & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* [Quality of

- education in schools: School internal and external conditions for mathematical, technical and overall competencies*] (pp. 65–82). *Zeitschrift für Pädagogik*, 45. Weinheim: Beltz.
- Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, 19(1–2), 221–240.
- Jäger, R. S. (2006). Diagnostischer Prozess. In F. Petermann & M. Eid (Eds.), *Handbuch der psychologischen Diagnostik [Handbook of psychological diagnostics]* (pp. 89–96). Göttingen: Hogrefe.
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73–84.
- Kärner, T., & Kögler, K. (2016). Emotional states during learning situations and students' self-regulation: process-oriented analysis of person-situation interactions in the vocational classroom. *Empirical Research in Vocational Education and Training*, 8(12), 1–23.
- Kärner, T., Fenzl, H., Warwas, J., & Schumann, S. (2019a). Digitale Systeme zur Unterstützung von Lehrpersonen—Eine kategoriengeleitete Sichtung generischer und anwendungsspezifischer Systemfunktionen [Technological support systems for teachers. A category-based review of generic and task-specific system functions]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 115(1), 39–65.
- Kärner, T., Sembill, D., Aßmann, C., Friederichs, E., & Carstensen, C. H. (2017). Analysis of person-situation interactions in educational settings via cross-classified multilevel longitudinal modelling: Illustrated with the example of students' stress experience. *Frontline Learning Research*, 5(1), 16–42.
- Kärner, T., Warwas, J., & Schumann, S. (2019). Addressing individual differences in the vocational classroom: towards a Teachers' Diagnostic Support System (TDSS). In T. Deißinger, U. Hauschildt, P. Gonon, & S. Fischer (Eds.), *Contemporary Apprenticeship Reforms and Reconfigurations. Proceedings of the 8th Research Conference of the International Network for Innovative Apprenticeships* (pp. 179–182). Zürich: Lit.
- Karst, K. (2012). *Kompetenzmodellierung des diagnostischen Urteils von Grundschullehrern [Competence-modelling of the diagnostic judgment of primary school teachers]*. Münster: Waxmann.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogenen Informationen als zielgeleiteter prozess. der lehrer als "flexibler denker" [Goal-directed processing of students' attributes: the teacher as "flexible thinker"]. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 175–186.
- Lau, N., Jamieson, G. A., & Skraaning Jr, G. (2012). *Situation awareness in process control: A fresh look*. Paper presented at the 8th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation and Control and Human-Machine Interface Technologies (NPIC and HMIT, July 22–26), San Diego, CA.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Cambridge: Elsevier.
- Lorenz, C., & Artelt, C. (2009). Fachspezifität und stabilität diagnostischer kompetenz von grundschullehrkräften in den fächern deutsch und mathematik [Domain specificity and stability of diagnostic competence among primary school teachers in the school subjects of german and mathematics]. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 211–222.
- Machts, N., Kaiser, J., Schmidt, F. T., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, 19, 85–103.
- Maqtary, N., Mohsen, A., & Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, 24(2), 169–190.
- McKenney, S., & Reeves, T. C. (2014). Educational design research. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 388–396). New York, NY: Springer.
- McKenney, S. E., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). London: Routledge.
- Mischel, W. (1968). *Personality and assessment*. London: Wiley.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.
- Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality*, 21(6), 789–810.
- Nielsen, J. (1993). Interactive user-interface design. *IEEE Computer*, 26(11), 32–41.
- Norman, G. (2009). Dual processing and diagnostic errors. *Advances in Health Sciences Education*, 14(1), 37–49.

- Power, D. J. (2002). *Decision support systems: Concepts and resources for managers*. Westport, CT: Quorum Books.
- Practorius, A.-K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction*, 52, 148–160.
- Rausch, T. (2013). Wie sympathie und ähnlichkeit leistungsbeurteilungen beeinflussen kann [How sympathy and affinity affect performance ratings]. *Erziehung und Unterricht*, 163(9–10), 937–944.
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331–363.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3–4), 237–263.
- Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., et al. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, 9(4), 297–323.
- Salmon, P. M., Stanton, N. A., & Young, K. L. (2012). Situation awareness on the road: Review, theoretical and methodological issues, and future directions. *Theoretical Issues in Ergonomics Science*, 13(4), 472–492.
- Sandoval, W. A. (2013). Educational design research in the 21st century. In R. Luckin, S. Puntambekar, P. Goodyear, B. L. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 388–396). New York, NY: Routledge.
- Schmitt, M., & Hofmann, W. (2006). Situationsbezogene Diagnostik. In F. Petermann & M. Eid (Eds.), *Handbuch der Psychologischen Diagnostik [Handbook of psychological diagnostics]* (pp. 476–484). Göttingen: Hogrefe.
- Schneider, S. C. (1987). Information overload: Causes and consequences. *Human Systems Management*, 7(2), 143–153.
- Sembill, D., Seifried, J., & Dreyer, K. (2008). PDAs als erhebungsinstrument in der beruflichen lernforschung—ein neues wundermittel oder bewährter standard? Eine Replik auf Henning Pätzold [PDAs as a tool for data collection in adult education research—New panacea or approved standard? A reply to Henning Pätzold]. *Empirische Pädagogik*, 22(1), 64–77.
- Shirley, M. L., & Irving, K. E. (2015). Connected classroom technology facilitates multiple components of formative assessment practice. *Journal of Science Education and Technology*, 24(1), 56–68.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400.
- Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, 37(1), 137–148.
- Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L. J., Mukama, E., Warusavitarana, A., et al. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology and Society*, 19(3), 58–71.
- Spinath, B. (2005). Akkuratheit der einschätzung von schülermerkmalen durch lehrer und das konstrukt der diagnostischen kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95.
- Stang, J., & Urhahne, D. (2016). Wie gut schätzen lehrkräfte leistung, konzentration, arbeits- und sozialverhalten ihrer schülerinnen und schüler ein? Ein beitrag zur diagnostischen kompetenz von lehrkräften [How well do teachers rate their students' achievement, attention, work habits and social behaviour? A contribution to the diagnostic competence of teachers]. *Psychologie in Erziehung und Unterricht*, 63(3), 204–219.
- Stanton, N. A., Baber, C., Walker, G., Salmon, P., & Green, D. (2004). Towards a theory of agent-based systemic situational awareness. In D. A. Vincenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness and automation: Current research and trends* (pp. 83–87). Proceedings of the 2nd Human Performance, Situation Awareness and Automation Conference (HPSAII). Mahwah, NJ: Lawrence Erlbaum.
- Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2009). Genotype and phenotype schemata and their role in distributed situation awareness in collaborative systems. *Theoretical Issues in Ergonomics Science*, 10(1), 43–68.
- Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., et al. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics*, 49(12–13), 1288–1311.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.

- Südkamp, A., & Möller, J. (2009). Referenzgruppeneffekte im simulierten Klassenraum: direkte und indirekte Einschätzungen von Schülerleistungen [Reference-group effects in a simulated classroom: Direct and indirect assessments of student performance]. *Zeitschrift für Pädagogische Psychologie*, 23(3–4), 161–174.
- Südkamp, A., Praetorius, A.-K., & Spinath, B. (2017). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, 76, 204–213.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: classrooms in the context of reform. *Assessment in Education: Principles, Policy and Practice*, 17(4), 399–417.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Urhahne, D., Timm, O., Zhu, M., & Tang, M. (2013). Sind unterschätzte Schüler weniger leistungsmotiviert als überschätzte Schüler? [Are underestimated students less achievement motivated than overestimated students?]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45(1), 34–43.
- Urhahne, D., & Zhu, M. (2015). Accuracy of teachers' judgments of students' subjective well-being. *Learning and Individual Differences*, 43, 226–232.
- Van den Berg, M., Harskamp, E. G., & Suhre, C. J. M. (2016). Developing classroom formative assessment in Dutch primary mathematics education. *Educational Studies*, 42(4), 305–322.
- Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. (2015). Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy and Practice*, 22(3), 324–343.
- Van Ophuysen, S., & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf – Anmerkungen zum Themenheft “Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz” [The quality of pedagogical diagnostics in the teaching profession—Annotations to the special issue “Teachers' diagnostic competences and their practical relevance”]. *Journal for Educational Research Online*, 7(2), 82–98.
- Van Ophuysen, S., & Lintorf, K. (2013). Pädagogische Diagnostik im Schulalltag. In S.-I. Beutel, W. Bos, & R. Porsch (Eds.), *Lernen in Vielfalt. Chance und Herausforderung für Schul- und Unterrichtsentwicklung* [Learning in diversity. Opportunity and challenge for educational development] (pp. 55–76). Münster: Waxmann.
- Warwas, J., Kärner, T., & Golyszny, K. (2015). Diagnostische Sensibilität von Lehrpersonen im Berufsschulunterricht: Explorative Prozessanalysen mittels Continuous-State-Sampling [Diagnostic sensitivity of teachers in vocational classes: Explorative process analyses using Continuous-State-Sampling]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(3), 437–454.
- Weinert, F. E., & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. Wagner, & B. Wolf (Eds.), *Schülergerechte Diagnose. Theoretische und empirische Beiträge zur Pädagogischen Diagnostik* [Student-friendly diagnostics. Theoretical and empirical contributions to pedagogical diagnostics] (pp. 11–29). Weinheim: Beltz.

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