peDOCS
Open Access Erziehungswissenschaften

DIPF

Klein, Martin; Wagner, Kai; Klopp, Eric; Stark, Robin

Fostering of applicable educational knowledge in student teachers: Effects of an error-based seminar concept and instructional support during testing on qualities of applicable knowledge

Journal for educational research online 9 (2017) 2, S. 88-114

Empfohlene Zitierung/ Suggested Citation:
Klein, Martin; Wagner, Kai; Klopp, Eric; Stark, Robin: Fostering of applicable educational knowledge in student teachers: Effects of an error-based seminar concept and instructional support during testing on qualities of applicable knowledge - In: Journal for educational research online 9 (2017) 2, S. 88-114 - URN: urn:nbn:de:0111-pedocs-148981

in Kooperation mit / in cooperation with:

WAXMANN
VERLAG GMBH
Münster · New York · München · Berlin

http://www.waxmann.com

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der Leibniz-Gemeinschaft

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 9 (2017), No. 2, 88–114
© 2017 Waxmann

Martin Klein, Kai Wagner, Eric Klopp & Robin Stark

Fostering of applicable educational knowledge in student teachers: Effects of an error-based seminar concept and instructional support during testing on qualities of applicable knowledge

Abstract

The project "Theorizing in Practice" uses typical theory application errors to foster applicable educational knowledge for the theory-based explanation of complex school situations. Earlier studies showed positive effects of error-based learning environments. However, a seminar concept expanding the learning environment by adding theoretical reflections of biographical learning and teaching experiences yielded only minor additional learning effects. The results were ascribed to insufficient sensitivity of the existing analyses towards qualitative changes in the learners' knowledge base. Hence, the present study examines to what extent enhanced analysis methods can assess qualities of knowledge based on a taxonomy by de Jong and Ferguson-Hessler (1996). The seminar concept (integrated/regular) and access to instructional support during testing (with/without) were varied experimentally. The participants' explanations in two test scenarios were analyzed with regard to structure, automation level and elaboration level of educational knowledge. As expected, structure and automation in the post test explanations in the integrated seminar were superior to those in the regular seminar. The structure level was similar to the pre-test, but performance (time-on-test) improved. Elaboration level of knowledge was superior in the integrated seminar. Access to instructional support during testing also improved structure and elaboration level, however, there was no significant interaction.

Keywords

Theory application, Evidence based practice, Learning from errors, Teacher education

---

Dr. Martin Klein (corresponding author) · Dr. Kai Wagner · Dr. Eric Klopp · Prof. Dr. Robin Stark, Department of Education, Saarland University, Campus, 66123 Saarbrücken, Germany
e-mail: martin.klein@mx.uni-saarland.de
k.wagner@mx.uni-saarland.de
e.klopp@mx.uni-saarland.de
r.stark@mx.uni-saarland.de

Förderung anwendbaren bildungswissenschaftlichen Wissens bei Lehramtsstudierenden: Effekte eines fehlerbasierten Seminarkonzepts und instruktionaler Hilfen während des Testens auf Qualitäten anwendbaren Wissens

Zusammenfassung

Das Projekt „Theoretisieren für die Praxis" fördert anwendbares bildungswissenschaftliches Wissen zur Erklärung komplexer schulischer Situationen mittels typischer Theorieanwendungsfehler. Frühere Studien zeigten positive Effekte fehlerbasierter Trainings. Ein diese Trainings erweiterndes Seminarkonzept zur theoretischen Reflexion biografischer Lehr-Lernerlebnisse erbrachte jedoch nur geringe Zusatzeffekte. Dies wurde mit der unzureichenden Sensibilität der verwendeten Analysemethoden in Bezug auf qualitative Veränderungen in der Wissensbasis der Lernenden begründet. Daher untersucht die vorliegende Studie, inwiefern verbesserte Analysemethoden Qualitäten von Wissen basierend auf einer Taxonomie von de Jong and Ferguson-Hessler (1996) erfassen können. Das Seminarkonzept (integriert (IS)/regulär (RS)) und die Verfügbarkeit instruktionaler Hilfen während der Testphase (mit/ohne) wurden experimentell variiert. Die Erklärungen der Teilnehmer in zwei Testszenarien wurden bezüglich ihres Strukturierungs- und Automationsgrades sowie der Verarbeitungstiefe bildungswissenschaftlichen Wissens ausgewertet. Wie erwartet waren diese in den Nachtest-Erklärungen im IS höher als im RS. Der Strukturierungsgrad war vergleichbar zum Vortest, jedoch verbesserte sich die Performanz (Bearbeitungszeit). Die Verarbeitungstiefe des Wissens war im IS höher. Die Verfügbarkeit instruktionaler Hilfen verbesserte Strukturierungsgrad und Verarbeitungstiefe, eine signifikante Interaktion blieb jedoch aus.

Schlagworte

Theorieanwendung, evidenzbasierte Praxis, Lernen aus Fehlern, Lehrerbildung

1. Problem description and aims

The application of scientific educational knowledge to improve pedagogical practice is gaining importance against the backdrop of the current debate on the introduction of evidence-based practice in the education sector (Bromme, Prenzel, & Jäger, 2014; Slavin, 2002, 2008). Evidence based practice refers to the consideration of the most recent scientific evidence for professional decisions and actions (Bauer, Prenzel, & Renkl, 2015, cf. Evidence-based Medicine Working Group, 1992). The German *standards for teaching and teacher education* (Standing Conference of the Ministers of Education and Cultural Affairs [KMK], 2004, 2015) consequently demand future teachers to be able to interpret, reflect on and make use of educational research findings for their classroom practice. Empirical data, however, require critical interpretation on a theoretical basis to gain the status of evidence (Bromme et al., 2014). Thus, future teachers also have to be enabled to make use of theories (Meier, 2006). This includes the theory-based explanation of complex school situations (cf. *retrospective understanding*; Beck & Krapp, 2006) as a form of scientific argumentation. Since both theories and evidence can be enlisted in this task, this article subsumes both categories under the term *scientific educational knowledge*.

Various problems occur due to the complexity of scientific knowledge application as well as the lack of its systematic training (Ohlsson, 1992; Bainbridge, 2011). Scientific educational knowledge remains *inert* (Renkl, 2006): Students are able to reproduce it in exams, but are unable to apply it to specific pedagogical situations. Even experienced teachers have problems understanding empirical data and show deficits in the argumentative use of evidence (Stark, Herzmann, & Krause, 2010; Wenglein, Bauer, Heininger, & Prenzel, 2015). Hence, evidence from educational research is hardly used as a resource for classroom practice or reflection by teachers (Hargreaves, 2000; Hetmanek et al., 2015; Neuweg, 2007). In addition, their scientific argumentation is prone to numerous errors (Stark, 2005) such as explanations of complex school situations on the basis of every day knowledge or subjective theories (Groeben, Wahl, Schlee, & Scheele, 1988) or faulty use of empirical evidence (Kuhn, 2010).

These errors, however, constitute learning opportunities. Based on earlier studies presenting typical scientific knowledge application errors in *integrated learning environments* (Reinmann & Mandl, 2006) to foster the acquisition of applicable educational knowledge (Wagner, Klein, Klopp, & Stark, 2014a, 2014b), we developed an error-based *integrated seminar concept*. It is based on a multitude of established instructional design principles and focuses on the theoretical reconstruction and reflection of biographical learning and teaching experiences (cf. KMK, 2004; 2014; Le Cornu & Ewing, 2008). The integrated seminar concept was compared to a regular teacher education seminar on theory-based reflection in the present study. Taking into account that future teachers will have access to external resources for the evaluation of school situations or course planning (Hetmanek et al., 2015), we also examined the effects of an instructional support measure during the learning tests to increase the ecological validity of the testing situation. We assessed the effects on different properties of *applicable knowledge* according to Krause (2007) in the theory-based explanation of complex school situations, drawing on a knowledge taxonomy by de Jong and Ferguson-Hessler (1996).

2. Theoretical background

2.1 Theory-based reflection of pedagogical experiences as a form of theory application

The ability to apply scientific educational knowledge to perform a theory-based reflection of pedagogical experiences or questions is a part of a teacher's competences as outlined in the *standards for teaching and teacher education* (KMK, 2004; 2015). This particular application of theoretical knowledge is described in Beck and Krapp's (2006) model of four basic forms of theory application: *Goal attainment* (technology: how can one attain a certain goal?), *impact assessment* (prognosis: what consequence B will phenomenon or intervention A have); *retrospective understanding* (explanation: How did B come to pass?); and *differentiated perception* (which situational information is relevant?).

While Beck and Krapp (2006) place *goal attainment* as the starting point, theory-based reflection begins with a given situation, which subsequently has to be reconstructed with regard to relevant information (*differentiated perception*) and then explained (*retrospective understanding*). Interventions or prognoses can then be derived from that explanation. This reflection is deliberately removed from the actual situation to enable students to evaluate their assessments and compare them to those derived from scientific educational knowledge (cf. Neuweg, 2007). With practice, this should sharpen the students' observation of pedagogical situations and enable them to make theory-backed decisions under (time) pressure (which in turn can then be reflected upon, and so on).

2.2 Learning from errors in integrated learning environments

Learning from advocatory errors (Oser, 2007) is based on presenting typical errors in a domain and contrasting those with the correct solutions (cf. Durkin & Rittle-Johnson, 2012; Stark, Kopp, & Fischer, 2011). As a result, learners acquire *negative knowledge* (negative conceptual knowledge: how something *is not*, negative procedural knowledge: how something *is not done* and negative strategic knowledge: which *strategies are not working*) through self-explanation processes (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). Negative knowledge prevents learners from future errors, since erroneous concepts, actions and strategies can be consciously excluded (Oser, 2007). However, this requires a detailed error analysis to ensure that learners understand clearly what is wrong and why something is wrong (Curry, 2004). Additionally, in learning from advocatory errors, learners have to adopt the perspective of a fictitious protagonist committing the errors.

Earlier studies successfully realized these requirements in *integrated learning environments* in the domains of medicine (Stark et al., 2011) and education

(Wagner et al., 2014a, 2014b). Integrated learning environments combine relevant, authentic problems with instructional support measures such as prompts or feedback (e.g. Durkin & Rittle-Johnson, 2012; Stark et al., 2011) or worked examples that present the problem and its solution as well as each step towards it (Renkl & Atkinson, 2010). Problem-based learning requires learners to actively construct knowledge to solve a problem (Reinmann & Mandl, 2006). This results in deeper understanding of concepts and more stable learning outcomes (Greeno, Collins, & Resnick, 1996; Renkl, 2014). Instructional support is necessary, however, to ensure that low prior knowledge learners profit as well (Kirschner, Sweller, & Clark, 2006).

A follow-up study (Klein, Wagner, Klopp, & Stark, 2015) examined the effects of a seminar concept based on *blended learning* and *fading of instructional support* as an integration of problem- and instruction-based learning on a seminar level. *Blended learning* refers to a combination of classroom teaching and self-regulated learning. The didactic intertwining of different methodical and medial approaches (Steffens & Reiß, 2009) can enhance the quality and efficiency of learning by highlighting different aspects of a learning subject (cf. Kerres & Jechle, 2002). However, just as in problem-based learning, learners require instructional support such as clear instructions and access to external resources (e.g. feedback or support materials from teachers, supported transfer phases, cf. Mandl & Kopp, 2006) during the self-regulated learning phases. While instructional support is especially helpful for learners with low prior knowledge (Reinmann & Mandl, 2006; Renkl, 2014; Schmidt, Loyens, Van Gog, & Paas, 2007), its advantages decrease over the course of the learning process (Eiriksdottir & Catrambone, 2011; Salden, Aleven, Schwonke, & Renkl, 2010). It can even have potentially detrimental effects on learning, e.g. when instructional support interferes with a learner's previously applied problem-solving strategies (*expertise-reversal*-effect, cf. Salden et al., 2010). By *fading of instructional support* (Wecker, 2012; Wecker & Fischer, 2011) the intensity and amount of instructional assistance are continually reduced to adapt to the learning progress (Renkl & Atkinson, 2010). Salden et al. (2010) point out the importance of clearly incremented levels within the fading process. Eiriksdottir and Catrambone (2011) discuss different forms of *fading* procedures, concluding that detailed, closely task-related instructions foster initial learning, whereas instructions with a higher degree of abstraction induce problem solving and foster reflection and transfer.

Despite these didactic measures, the seminar concept only yielded minor additional learning outcomes, possibly due to insufficient sensitivity of the analysis methods to *qualitative* changes in the learners‘ knowledge base (Klein et al., 2015). These shortcomings were addressed in the present study by enhanced analysis methods and an additional instructional support measure.

2.3 Instructional support during learning tests

Teachers usually have access to various external information resources ranging from collegial feedback to educational knowledge from specialized journals to solve problems (Hetmanek et al., 2015). Hence, instructional materials providing crucial information on a problem-solving task such as a test increase the authenticity of that task. Studies on *open book testing* as a form of instructional support during testing show not only a preference of students towards tests in which learning materials are provided as a resource, but also positive effects on knowledge application tasks (Agarwal & Roedinger, 2011; Brightwell, Daniel, & Stewart, 2004; Theophilides & Dionysiou, 1996). Access to conceptual knowledge shifts both the learning and testing focus from knowledge recall to knowledge *application*, such as reasoning or problem solving (Mekala, 2011).

However, this requires the tests to present actual knowledge application tasks instead of recall tasks (Agarwal & Roedinger, 2011; Ioannidou, 1997). Additionally, even with knowledge access in the test, the learning outcomes are not independent from the learners' preparation level (Thorndike, Cunningham, Thorndike, & Hagen, 1991). Providing all learning materials during a test can entice learners to spend much time searching or verifying information. Well-prepared learners retain an advantage, since their search is more specific (ibid.). Knowing that there will be access to all relevant information can also reduce preparation time, decreasing long-term learning effects (Agarwal & Roedinger, 2011) and impairing the development of an elaborate conceptual knowledge base, an essential component of applicable knowledge (Krause, 2007).

From these considerations, we inferred three requirements of effective instructional support during testing: (1) Its content has to be reduced to the most relevant information for the task at hand to optimize the search for information (Thorndike et al., 1991); (2) it has to build upon knowledge acquired in the learning phase (Krause, 2007); and (3) the test itself has to assess knowledge *applications* such as reasoning or problem solving (Ioannidou, 1997; Agarwal & Roedinger, 2011). A further question that has not been addressed yet is the effect of providing *negative knowledge* (Oser, 2007; see section 2.2) during a test. From a theoretical perspective, negative knowledge about errors and error avoidance strategies should safeguard learners against erroneous decisions or actions and thus be a valuable resource in a test.

2.4 Types and qualities of knowledge

Krause (2007) defines applicability as a property of knowledge that requires the successful interaction of different knowledge dimensions to solve problems. The present study's understanding of these knowledge dimension draws upon a two-dimensional taxonomy by de Jong and Ferguson-Hessler (1996), which differentiates between *types* and *qualities* of knowledge. Analogous to Oser (2007), de

Jong and Ferguson-Hessler's knowledge *types* refer to the way knowledge is categorized (*conceptual:* singular concepts and facts; *procedural:* concrete actions; *strategic:* strategies and action sequences). Even though the terms are similar, de Jong and Ferguson-Hessler's taxonomy differs in some aspects from others such as the *ACT-R-Theory* (Anderson et al., 2004). The *ACT-R-Theory* states that declarative (i.e. conceptual) knowledge *about* actions becomes procedural knowledge through a process called *compilation*, wherein an individual develops *production rules* that contain information on the execution and sequence of actions. De Jong and Ferguson-Hessler instead regard conceptual and procedural knowledge as different knowledge *types*, which can each have a *degree* of compilation that ranges from declarative (unconnected pieces of information or step-wise operations) to *compiled* (holistic, intuitive understanding or routine execution) on an additional dimension they refer to as *automation level* of knowledge. These additional dimensions, the *qualities of knowledge,* are in the focus of this study. They also include the *structure* of knowledge and its *level of depth* (referred to in this article as *elaboration level*). Highly *structured* knowledge is hierarchically and meaningfully organized and facilitates the generation of interrelations and coherent units of meaning such as schemata (Rumelhart, 1980). Such knowledge structures are associated with expertise in a domain (e.g., medicine, see Boshuizen & Schmidt, 1992) and allow for faster and less demanding (in terms of cognitive resources) access to information. This increases the aforementioned *automation level* as well as the applicability of knowledge in real world situations (Stark, 2001), since problem solving processes can be sped up by easier knowledge retrieval and the availability of additional cognitive resources for problem solving (de Jong & Ferguson-Hessler, 1996; cf. *efficiency*, Eysink et al., 2009). Knowledge with a high *elaboration level* (cf. *deep-level knowledge* as opposed to *surface-level knowledge*; de Jong & Ferguson-Hessler, 1996) is characterized by a high level of abstraction of basic principles, concepts and procedures in a domain, which facilitates evaluation and critical thinking (Marton & Säljo, 1976). *Structure* and *elaboration* level of knowledge are interdependent as well: The construction of functional schemata and knowledge hierarchies requires abstracted knowledge (de Jong & Ferguson-Hessler, 1996). This theoretical interdependency indicates that they should be positively correlated. Since the qualities can be assessed for each knowledge type, they allow for the detection of changes in a learner's knowledge base that may be obscured by focusing solely on changes in the knowledge types.

3. Research questions and hypotheses

Based on our theoretical considerations, an error-based integrated seminar concept didactically implemented by *blended learning* and *fading of instructional support* was developed to foster three *qualities* of applicable scientific knowledge for the explanation of complex school situations. The seminar concept focused on the theoretical reconstruction and reflection of biographical learning and teaching experiences (cf. KMK, 2004, 2014; Le Cornu & Ewing, 2008). This included the analysis, identification and avoidance of argumentation errors and the autonomous research and application of appropriate theories in a scientific explanation. An instructional support measure was added to the concluding learning tests by providing students with negative knowledge about argumentation errors to support the students in the critical evaluation and choice of theories and empirical data (cf. Wenglein et al., 2015) and increase the ecological validity of the test situation. A regular seminar concept that included no references to argumentation errors and did not employ the integrated seminar's didactic measures was used as a control group. The following research question and hypotheses were formulated:

To what extent do the integrated seminar concept and the instructional support during testing foster structure, automation level *and* elaboration level *of applicable scientific knowledge*?

1. *Main effect of the seminar concept:* Participants in the integrated seminar concept were expected to show better *structure of knowledge* in their explanations than the control group and improve from pre- to post-test. Consequently, the *automation level* should be higher in the post-than in the pre-test due to easier knowledge retrieval. The *elaboration level* of knowledge in the post-test should exceed that of the control group, since the integrated seminar concept also fostered evaluation and critical thinking with regard to theory selection.

2. *Main effect of instructional support during testing:* Providing negative knowledge during testing should improve the *structure of knowledge* and *elaboration level* of the students' explanations in the post-test by supporting them in the exclusion of argumentation errors and in theory selection.

3. *Interaction between seminar concept and instructional support during testing*. We expect the effects of the integrated seminar concept on the *structure* and *elaboration level* of knowledge to be bolstered by the availability of negative knowledge since it enables the participants to focus on knowledge application instead of recall in the exclusion of argumentation errors and in theory selection.

4. Method

4.1 Sample and design

N = 135 student teachers (72f, 13 not indicated) were recruited from obligatory seminars in teacher education. The mean age was 22.2 years (*SD* = 3.77). The seminars consisted of two phases (see table 3; cf. Klein et al., 2015): In seminar phase 1 (SP1), *all* participants were trained with the learning environment. The experimental variations were implemented only in seminar phase 2 (SP2). At the beginning of SP2, the seminars receiving the integrated seminar concept were randomly selected. Structure and automation level were assessed at the beginning and end of

SP2 by an *application task* (details in section 4.5), resulting in a 1x2-factorial design with repeated measurements (*seminar concept*: integrated vs. regular, see table 1). An additional *transfer task* (details in section 4.5), was administered only at the end of SP2 to assess structure and elaboration level. In this test, the availability of instructional support during testing was varied within a 2x2-factorial design (*seminar concept* x *instructional support*: with vs. without, see table 2).

Table 1: Design A: Experimental design with repeated measurements design in the application task

Seminar concept	
integrated	regular
N =67 40f, 6 n.i., M_{age}= 23.4 (*SD* = 3.47)	*N* =68 40f, 7 n.i., M_{age}= 22.9 (*SD* = 3.67)

Note. f = female, n.i. = not indicated

Table 2: Design B: Experimental design in the transfer task

		Seminar concept	
		integrated	regular
Instructional support	with	*n* = 35 17f, 3 n.i., M_{age}= 22.9 (*SD* = 2.53)	*n* = 30 22f, 5 n.i., M_{age} = 22.8 (*SD* = 4.22)
	without	*n* = 28 17f, 1 n.i., M_{age} = 23.9 (*SD* = 4.37	*n* = 29 16f, 3 n.i., M_{age} = 23.3 (*SD* = 3.16)

Note. f = female; n.i. = not indicated. Different sample sizes due to participants not indicating whether or not they had received instructional support. These participants were excluded from the analyses.

4.2 Procedure

After the pretest (t1) in the first seminar session, *all* participants worked with the learning environment in SP1 (see table 3). Participants studied four case-based worked examples (Renkl & Atkinson, 2010) of school problem scenarios. A fictional young teacher acts in place of the participants and formulates erroneous explanations of problematic situations in his class (see figure 1 and 2). A school psychologist points out his errors in a detailed error analysis and presents a theory-based, scientific explanation (see figure 3) as a contrast as well as his strategies to avoid the respective errors.

Figure 1: Example of a school problem situation

1. The situation

Roman (27) has recently graduated and started teaching at a the local high school. A few weeks into the school year, he witnesses the following situation upon entering the classroom:

Markus, a tall and burly student who is quite popular in class, is trying to throw the backpack of his classmate Peter out of the window of the second floor. Peter is trying to oppose him and they grapple. Peter is no match for Markus, who is dealing hard blows and kicks. Their classmates do not interfere, but while some of them pretend not to notice, others cheer Markus on. Just as Roman is about to step in, Jenny, one of the students, seizes the initiative. She steps between the opponents and tries to prevent further escalation. With her help, Roman is able to stop the fight. Peter is consoled, Markus is sent to detention and Roman believes the issue to be over for now.

But when Roman thinks back, he realizes that there have been numerous encounters between the two students, always ending with Peter in tears. His colleague, Ms Schmidt, reports another scuffle just the other day. Roman wonders whether there might be more to these incidents than a normal dispute between students.

In the afternoon, he tries to explain the situation based on what he has witnessed and what he remembers from his studies.

Figure 2: Example of an erroneous explanation

2. Romans explanation

The next day, Roman presents his explanation to you in the teachers‘ lounge. He has even prepared a handout, because he wants to make sure that his explanation is scientifically viable. The school psychologist is there as well and overhears it.

Please read Roman‘s handout attentively!

Explanation of the incident between (the aggressor) and Peter (the victim):

- Since Markus has attacked Peter repeatedly, he seems to be a rather aggressive person. From my studies I remember that this might mean that aggressiveness is a personality trait of Markus. Traits don‘t change over time, so Markus‘ inherent aggressiveness might explain why he keeps on attacking other students.

- According to the Frustration-Aggression-Hypothesis (Dollard et al, 1941) that I also remember from my studies, aggression is caused by frustration. Frustration occurs when a person is prevented from performing an activity that they would like to do. So Peter seems to be hindering or frustrating Markus in some way, which explains why Markus attacks him.

At this point, the school psychologist, Mr Jung, asks to see the handout, since the incident pertains his responsibilities as well. Immediately, he realizes that Roman has made 3 substantial errors in his explanation. To help Roman, he walks him through each aspect of the explanation and explains where Roman is mistaken.

Please read Mr Jung‘s comments attentively!

Figure 3: Example for the school psychologist's error analysis

3. The school psychologist's error analysis (3/3)

Romans error: „Reference to unsuitable theories“

The error „reference to unsuitable theories“ is made when an explanation is based on an empirically well-established theory, but one that does not fit the situation. Thus, the theory is unsuitable to explain the situation.

Romans explanation

Frustration-Aggression-Hypothesis (Dollard et al, 1941):

- According to the Frustration-Aggression-Hypothesis (Dollard et al, 1941) that I also remember from my studies, aggression is caused by frustration. Frustration occurs when a person is prevented from performing an activity that they would like to do. So Peter seems to be hindering or frustrating Markus in some way, which explains why Markus attacks him..

School psychologist's feedback

1. **What is wrong and why is it wrong?**
   This error occurs if you do not check whether or not the conditions (here, of a certain behavior) named in the theory are present in the situation. As a consequence, it is not possible which situational factors caused the phenomenon in question or the wrong factors are identified as causal.

2. **Situational context**
   Although the *Frustration-Aggression-Hypothesis* is an empirically well-established theory, it does not fit the situation. There is no frustration apparent in Markus' behavior, which rules it out as a probable cause for his aggression.

The participants' task was to read the worked examples and answer two prompts regarding error definitions and error avoidance. All errors presented in the erroneous explanations were examples of frequent errors in scientific argumentation tasks (Stark, 2005). E.g., a fundamental attribution error (Gilbert & Malone, 1995), is a frequent error in the explanation of discipline problems, plausible in this context and would adversely affect the planning of an appropriate intervention. In the fourth session, the first post-test (t2) was administered. The two lecturers giving the seminars received a manual for SP1 to ensure standardization. The same lecturers were responsible for SP2.

SP1 served to establish a comparable knowledge base with regard to theory-based explanation and argumentation errors for all participants, which was ensured by the first post test at t2 (see section 4.5 for details). At the beginning of SP2, the participants were assigned to the integrated and regular seminar concepts and given work assignments in the eight sessions of phase 2 (variation A). In the second post-test (t3), which was administered in the last session, participants were randomly assigned to the groups with and without instructional support (variation B, see table 3). There were no homework assignments, students engaged with the seminar content only during the seminar sessions (not controlling for voluntary additional work).

Table 3: Seminar procedure and experimental groups

Phase 1: learning environment, 4 weeks			**Phase 2:** work assignments, 8 weeks	
Session 1	**Session 2-3**	**Session 4**	**Session 5-12**	**Session 13**
Pretest (t1) sociodemo-graphic data structure of knowledge	**Training with the learning envi-ronment** 4 school problem scenarios regard-ing: – bullying – group phenomena – achievement motivation – discipline problems	**Post-test 1 (t2)** structure of knowledge (appli-cation; base-line for phase 2) automation level	**Integrated seminar concept** step-wise construc-tion of a school problem scenario **Regular seminar concept** standardized work assignments regarding theory application	**Post-test 2 (t3)** structure of knowledge automation level elaboration level **with/without** instructional support

Note: For detailed information on the tests see section 4.5

4.3 Experimental variation A: integrated vs. regular seminar concept

Table 4 provides a short comparison of the two seminar concepts.

Table 4: Juxtaposition of the seminar concepts

	Integrated seminar concept	**Regular seminar concept**
Topic	Theory-based reflection of own experiences/working on own questions	Theory-based reflection of own experiences/working on own questions
Primary didactic orientation	*Error-based*: Task includes – erroneous explanation – detailed error analysis – context-specific error avoidance strategies	No error orientation (correct explanation only)
Feedback characteristics	Referral to typical errors in feedback on students‘ argumentations Provision of suitable avoidance strategies	Suggestions on possible improvements in students‘ argumentation
Additional didactic measures	*Blended learning:* lecturer-centered introduction phase + self-regulated working phase *Fading of instructional support:* systematic reduction of input, from task-related, detailed instructions and examples to abstract outlines	Self-regulated learning only Minimal, but constant instructional support

Integrated seminar concept: The integrated seminar concept referred systematically to the structure and concept of the learning environment in SP1 by having participants construct school problem scenarios based on their own experiences or questions. This included a problem description, an erroneous explanation, a detailed error analysis, and context-specific error avoidance strategies. The latter would subsequently be used to avoid errors in a scientific explanation based on theories selected by the participants. Thus, applicable knowledge acquired in the training with the learning environment in SP1 was connected to the participants' own experiences to improve relevance and authenticity (Reinmann & Mandl, 2006) and foster reflection and transfer (*retrospective understanding;* Beck & Krapp, 2006). Problem- and instruction-based learning were integrated by combining *blended learning* with *fading of instructional support*. *Blended learning* was implemented by dividing six of the eight seminar sessions into a lecturer-centered introduction phase and a self-regulated working phase. In the introductory phase (approx. 15 min), the lecturers provided work assignments as well as detailed instructions and examples for each session. For the remainder of each session, participants worked on the assignments. The lecturers were available for questions or feedback during this time. In two sessions (8 and 12), the lecturers gave feedback on remaining errors in the work assignments and provided suitable avoidance strategies, reinforcing the connection to SP1 (see figure 4).

Fading of instructional support was implemented in two steps over the whole duration of SP2. Instructional support was systematically reduced: in the first three sessions dealing with tasks such as *writing a draft of a complex school situation* (which were not part of the training with the learning environment in SP1), closely task-related, detailed instructions and examples were provided. In session 9-11 (see figure 4), instructional support was reduced to providing an outline for the respective task (e.g., error analysis).

Figure 4: Blended learning and fading of instructional support in seminar phase 2

Session	Lecturer task	Student task
5	**Asssistance** in the formulation of a problem description	Draft of a school problem scenario
6	**Asssistance** in the formulation of an erroneous explanation integrating instructional errors from SP1	Draft of an erroneous explanation
7	**Information** on source analysis and differences between everyday and scientific educational knowledge	Theory research and summarization
8	**Feedback** on current progress	Presentation and group discussion/feedback
9	**Structural outline** of error analysis provided based on learning environment in SP1	Error analysis
10	**Structural outline** of error avoidance strategies provided based on learning environment in SP1	Formulation of error avoidance strategies with regard to the situational context
11	**Structural outline** of a scientific explanation provided based on learning environment in SP1	Formulation of a theory-based explanation
12	**Feedback** on completed scenarios	Presentation and group discussion/feedback

Fading: reduction of instructional support over time

Blended learning: Alternating lecturer-centered and self-regulated learning phases

Regular seminar concept: Participants in the regular seminar concept received work assignments on theory application. They were instructed to describe problematic school situations and to formulate a theory-based explanation on how they had come to pass, e.g., explaining discipline problems using social psychological theories about group phenomena. The participants chose their topics themselves and received elaborate feedback on their explanations. However, contrary to the integrated seminar concept, they were not instructed to write an erroneous explanation from the teacher's perspective and perform the corresponding error analysis, nor did the lecturers reference the errors presented in SP1 in the feedback on the participants' explanations. Instead, the feedback was limited to suggestions on how to improve the explanation, e.g., suggestions on useful or different theories. This way, participants in the regular seminar concept practiced theory-based explanations as well, however, the systematic referral to the training with the learning environment in SP1 was missing, as were the didactic principles of *blended learning* (participants only received written assignments and limited feedback) and *fading* (the rather minimal degree of instructional support was not reduced further over time). The number of sessions was identical to the integrated seminar concept.

4.4 Experimental variation B: instructional support during testing

Instructional support in the *transfer test* (see section 4.5) was implemented by a glossary referring to negative conceptual and strategic knowledge about argumentation errors which participants had acquired during the training with the learning environment in SP1 (see figure 5). All error definitions included in the learning environment as well as the corresponding error avoidance strategies were provided. In accordance with section 2.4, the instructional support measure was very specific to (1) optimize the search for a specific information and (2) link to the knowledge base from SP1. Participants were instructed to refer explicitly to the glossary to support their choices in the selection of theories and the formulation of their explanations by avoiding typical errors. Figure 5 shows an example from the glossary.

Figure 5: Instructional support during testing: Error definitions and error avoidance strategies

Fundamental attribution error	
Definition:	A tendency of a viewer to ascribe a person's observable behavior to their psychological disposition instead of situational factors or to overestimate the influence of a person's disposition and underestimate situational factors.
Error avoidance strategy:	Be aware of your own tendency to overestimate dispositional factors such as traits. Make sure to assess situational factors and consider them carefully in your explanation.

4.5 Instruments

The control variables *age* and *gender* were assessed by questionnaires. *Time-on-test* was assessed by having the participants note start- and end-time directly on the test sheets.

The three cognitive variables *structure, automation level* and *elaboration level* (de Jong & Ferguson-Hessler, 1996) of applicable knowledge according to Krause (2007) were operationalized as follows: Two tests were administered, in which students had to formulate explanations of a problematic school situation (presenting a phenomenon which was in any way detrimental to the teacher, the lesson delivery and/or one or more students) based on a given selection of theories (see figure 6).

Figure 6: Selection of theories provided in the learning materials (page 1 of 2)

5. The school psychologist‘s explanation: Theories

1. Bullying (definitory aspects according to Scheithauer, 2008):
- Special type of Aggression in school classes, as opposed to disputes
- Characterized by repeated one-sided incidents over a longer period (more than 3-6 months)
- Individual attacks, usually by one main aggressor
- Imbalance between aggressor and victim regarding physical strenght, social competence, social standing in favor of the aggressor
- Goal-driven subtype of aggression: direct & indirect damaging, exploitation of imbalance to increase own social standing → systematic abuse of power with proCharakter
- Victim can hardly avoid the attacks because of the group structure (e.g. school class)
- 3 types physical, verbal, relational bullying
- Classmates assume different roles: assistents (directly assist attacker, profit from increased standing), amplifier (verbally support attacker, without own stakes in the fight), bystander (passive support by non-intervention), defender (intervene to defend victim, usually on equal or higher social standing as attacker)

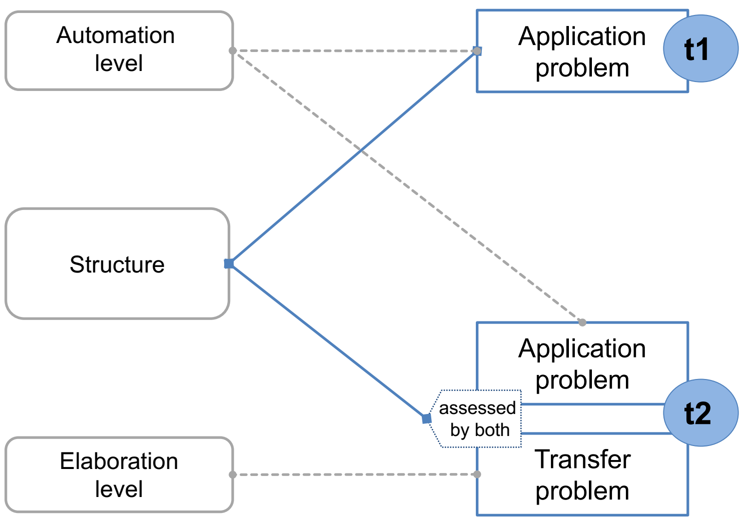
2. Instrumental aggression
- Socially inacceptable means of goal attainment
- Cost-benefit-analysis decides about use of aggression/violence, no direct hostility towards victim necessary
- Particularly associated with social goals such as increase of social standing or power or positive self portrayal
- Use of violence depends on a person‘s available resources, alternatives, learning history with regard to violence, but also on social factors (how far does the immediate environment condone or judge violent behavior)
- The need for positive self portrayal can foster aggression if the immediate environment assigns positive value to aggression

The theories had previously been selected by the authors and used to formulate a model scientific explanation that served as the basis for the *structure of knowledge* score (see below).

The students' explanations were then analyzed with regard to *structure, automation level* and *elaboration level* of knowledge. The first test was presented twice and presented a problem scenario on school bullying that was structurally similar to the first training scenario in SP1 (*application problem*; cf. Wagner et al., 2014b; see table 3).

The second test (*transfer problem*) presented a novel problem situation (a school student's tolerance of uncertainty (König & Dalbert, 2007) not fitting the teaching style) and an additional theory selection task (see below) and was only presented at the end of SP2 (t3). *Structure* of knowledge was assessed by both tests. Additionally, the *application problem* was used to assess the *automation level,* and the *transfer problem* was used to assess the *elaboration level* of knowledge (see figure 7).

Figure 7: The two tests and the dependent variables assessed by each test



Note: Lines denote which variable was assessed by which test.

Both tests were based on *key feature*-tests used in medical education, which assess whether essential steps and procedures were observed in the diagnosis and therapy of clinical problems (Kopp, Möltner, & Fischer, 2006). Analogously, participants had to observe the essential steps of a theory-based explanation. In accordance with section 2.4, the tests demanded *knowledge applications*. Based on the complexity of the tasks and time-on-test data from earlier studies, only two tests were administered.

Structure of knowledge: Structure of knowledge was assessed as an aggregate of two scores: a *content* score and a *quality of argumentation* score. To rate the *content of the explanation*, we examined the participants' explanations in both tests on whether they had identified relevant situational aspects and referred to scientific theories and models, or whether they had disregarded relevant information or overgeneralized and referred to subjective theories or individual experiences. Scores for *content of the explanation* were based on the comparison to a model explanation previously formulated by the authors as a scientifically accurate theory-based analysis with regard to the criteria mentioned above (the model explanation did not claim exclusivity, as different interpretations are possible in school situations). Points were awarded for each aspect in the students' explanation that was either found in the model explanation, could be used to append it or presented a different, but viable interpretation (application problem: max. 16, transfer problem: max. 20 (greater problem complexity due to more interacting variables). Additionally, the *quality of the argumentation* itself was assessed by an analysis oriented on Toulmin (1958) rating the coherence of argumentation as well as the quality of the connection to the backing. According to de Jong and Ferguson-Hessler (1995, 1996), a higher level of underlying *structure* of knowledge can be assumed if situational facts, theories, concepts and models are linked or hierarchically organized in a meaningful way to explain a situation. An additional maximum of four (4) points were awarded for the *quality of the argumentation*. The theoretical maximum for the sum score was 20 points for the application problem and 24 points for the transfer problem.

Automation level: The *automation level* was operationalized as task performance by recording time-on-task in the *application problem* and calculating a quotient from the time taken and the *structure* of knowledge scores. The quotients before and after SP2 (t2 and t3) were compared. A higher automation level was ascribed if participants scored equally or better at t3 than t2 in *less* time or took a similar amount of time to score substantially higher. This approach was based on suggestions by de Jong and Ferguson-Hessler (1995, 1996) for the assessment of automation. Because of the calculation method, the instructional support materials (error glossary) were not provided in the *application problem*, since the participants' referral to the materials would confound time-on-test data.

Elaboration level: In the *transfer problem*, participants had to explain a novel situation *and* choose from a selection of theories. This test was considered more complex than the application problem because of a higher number of interacting variables in the situation itself (persons, time information, problem concepts) and because some of the theories provided were inadequate for an explanation of this particular situation. Participants had to choose from six theories (three adequate, three inadequate theories). This requires evaluation and critical thinking, which are associated with the *elaboration level* of knowledge (de Jong & Ferguson-Hessler, 1995, 1996). Participants were instructed to write down their reasoning on this task. We also assessed whether they referred to negative knowledge (error avoidance strategies) in their *choice of theories*. This test was only administered at the end of SP2 (t3), since theory research and selection had not been part of the training with the learning environment, but were only fostered in SP2. One point was awarded for the exclusion of each inadequate theory with a correct justification (max. 3). One more point was awarded for referral to error avoidance strategies in the justification. Participants could score a maximum of 4 points in this task.

Scoring procedure: Both tests were scored by the two lecturers. The raters did not know which test belonged to which experimental condition. Scoring deviations were settled by discussion. This procedure was tested in earlier studies (cf. Klein et al., 2015; Wagner et al., 2014b). Interrater reliability (Cohen's κ) was .91 (t2) and .88 (t3) for the *structure* assessment in the application problem; .86 for the *structure* assessment and .87. for the assessment of the *elaboration level* in the transfer problem. Bivariate correlations using Pearson's *r* showed that the cognitive variables *structure*, *automation level* and *elaboration level* of knowledge correlated significantly (.23 < r< .79; see table 5).

Table 5: Correlations between the cognitive variables at t3

	Structure[b]	Elaboration Level[b]	Automation Level[a]
Structure[a]	.47**	.23*	.79**
Structure[b]		.57**	.36**
Elaboration level[b]			.24**

Note. [a]= application problem; [b]= transfer problem

* p < .05; **p < .001

4.6 Statistical analyses

Since all scores were ascribed interval scale level, the effect of the seminar concept on the *structure* of knowledge and on the *automation level* (quotient structure of knowledge/time-on-test) in the application problem was examined by a MANOVA. Subsequent analyses including the repeated measures factor were then performed by ANOVAs for each variable by a repeated measures 2x2-ANOVA. Effects of the seminar concept and the instructional support during testing on *structure* and *elaboration level* of knowledge in the transfer problem were calculated by a 2x2-MANOVA. Level of significance for the differential hypotheses was set at α = .05. Equivalence between the groups regarding the control variables was analyzed on an α-level of .20. Effect sizes are reported as η_p^2 *(0.01* = small, 0.06 = medium, 0.14 = strong effect; Cohen, 1988) for ANOVAs. Missing values were excluded list wise.

5. Results

5.1 Internal validity and preliminary analyses

For both experimental designs, there were no significant differences between the groups regarding age, gender, structure of knowledge, time-on-test and automation level at t2 (all p > 20).

5.2 Structure of knowledge and automation level in the application problem

All descriptive statistics are presented in tables 6 and 7. A significant large multivariate effect of the seminar concept on structure and automation level of knowledge in the application problem was confirmed by a MANOVA ($F(2,127) = 12.8$, $p < .001$, $\eta_p^2 = .17$). Subsequent analyses including the repeated measures factor were calculated by a 2x2 ANOVA with repeated measures.

Structure of knowledge. In the *application problem*, the seminar concept had a significant medium main effect on *structure* of knowledge, corresponding to our hypothesis ($F(1,134) = 12.4$, $p = .001$, $\eta_p^2 = .09$). Participants in the integrated seminar condition scored higher than those in the regular seminars condition (see table 6). There was also a significant small effect of the repeated measures factor ($F(1,134) = 6.25$, $p = .012$, $\eta_p^2 = .05$). Contrary to our hypotheses, however, the means of the test scores decreased from t2 to t3. Descriptively, the decrease was markedly smaller in the integrated seminar group. In an isolated analysis of this group, there was no significant effect of the repeated measures factor ($F < 1$). In the regular seminar concept group, the decrease was substantially larger and the effect was significant and large ($F(1, 67) = 11.2$, $p = .001$, $\eta_p^2 = .14$). In contrast to our expectations, there was no significant interaction between the group factor and the repeated measures factor ($F(1,134) = 3.19$, $p = .08$).

Automation level (quotient *structure of knowledge/time-on-test*): As hypothesized, there was a significant large main effect of both the seminar concept ($F(1, 128) = 22.3$, $p < .001$; $\eta_p^2 = .15$) and the repeated measures factor ($F(1, 128) = 38.6$, $p < .001$; $\eta_p^2 = .23$) on the automation level of knowledge in the application problem. There was also a significant medium interaction between the repeated measures factor and the seminar concept (F (1, 128) = 12.9; $p < .001$; $\eta_p^2 = .09$). This was due to a significant large effect of the repeated measures factor (t2 to t3, $F(1, 123) = 94.9$, $p < .001$, $\eta_p^2 = .44$) as well as a significant small group difference at t3 ($F(1, 123) = 4.32$, $p = .04$, $\eta_p^2 = .03$) for *time-on-test* in favor of the integrated seminar group (see table 6). However, there was no significant interaction for time-on-test between the seminar concept (group factor) and the repeated measures factor ($F(1, 123) = 1.61$, $p = .21$). Both groups improved with regard to time-on-test. However, only the integrated seminar group's actual task performance improved: While their structure scores did not differ significantly from t2 to t3 (see section 4.5), they took significantly less time to solve the problem at t3 than at t2, and significantly less time than the regular seminar group. The regular seminar groups also spent less time on the task at t3 than at t2, but their structure scores decreased significantly.

Table 6: Means and standard deviations in the application problem

Time of measurement	t2		t3	
Seminar concept	integrated	regular	integrated	regular
Structure of knowledge	13.3 (4.54)	12.4 (4.55)	13.0 (4.08)	10.6 (4.01)
Time-on-test (in minutes)	32.5 (8.25)	34.1 (6.87)	25.2 (13.4)	21.0 (9.02)
Automation level	0.41 (.15)	0.36 (.16)	0.62 (.26)	0.42 (.18)

Note. Structure of knowledge max. 20 points; automation level = structure score divided by time-on-test.

5.3 Structure of knowledge and elaboration level of knowledge in the transfer problem

In the *transfer problem*, medium multivariate effects were found for structure and elaboration level of knowledge by a 2x2-MANOVA (seminar concept: $F(2, 124) = 5.96$, $p = .003$, $\eta_p^2 = .09$; instructional support: $F(2, 124) = 7.76$, $p = .001$, $\eta_p^2 = .11$), while the interaction did not reach the level of statistical significance (seminar concept * instructional support: $F < 1$).

Structure of knowledge: The univariate analyses showed a significant medium main effect of the group factor in favor of the integrated seminars ($F(1, 125) = 8.60$, $p = .004$, $\eta_p^2 = .06$). As hypothesized, there was also a significant medium effect of instructional support during testing ($F(1, 125) = 14.4$, $p < .001$, $\eta_p^2 = .10$). Participants that were instructionally supported scored higher than participants without instructional support (see table 7). Contrary to our hypotheses, the interaction effect between seminar concept and instructional support was not significant ($F < 1$).

Elaboration level: In the transfer problem, significant medium main effects of the seminar concept ($F(1, 125) = 9.19$, $p = .003$; $\eta_p^2 = .07$) and the instructional support measure ($F(1, 130) = 7.82$, $p = .006$; $\eta_p^2 = .06$) were shown, corresponding to our hypotheses. Both students in the integrated seminar concept and students receiving instructional support performed significantly better at selecting appropriate theories and justifying their choices than students in the regular seminar and students without instructional support. Contrary to our hypotheses, however, the interaction effect was not significant ($F < 1$).

Table 7: Means and standard deviations in the transfer problem (t3)

Seminar concept	integrated		regular	
Instructional support	with	without	with	without
Structure of knowledge	14.9 (5.82)	11.9 (4.54)	12.7 (4.93)	9.10 (3.69)
Elaboration level	2.21 (1.33)	1.52 (1.21)	1.49 (1.18)	0.97 (0.97)

Note. Structure of knowledge max. 24 points; Elaboration level max. 4 points

6. Discussion

6.1 Fostering of applicable scientific knowledge for theory-based reflection

The present study showed that applicable educational knowledge can be fostered by training students in the theory-based reflection and explanation of complex school situations. The error-based integrated seminar concept addressed typical problems and errors student teachers show in the application of educational theories (Stark, 2005) and demonstrated the usefulness of scientific knowledge as a resource (Hetmanek et al., 2015). As expected, the systematic connection of the knowledge acquired in SP1 with the participants' own experiences and questions (KMK, 2004, 2016; Nausner, 2010; Neuweg, 2007) and the ensuing active knowledge construction (Greeno et al., 1996; Renkl, 2014; Spychiger, 2004) resulted in an improvement of the participants' knowledge base on several dimensions. Additionally, the effectiveness of instructional support materials providing negative knowledge about typical errors was shown. Even though the additional provision of error-related materials did not increase the effect of the seminar concept further, the main effect of this factor shows that providing students with negative knowledge can indeed serve as a means of protecting students from typical errors (Oser, 2007). Regarding the current trend towards a more evidence-based practice in education, this intervention should serve to acquaint student teachers more closely with the scientific knowledge base of their profession (Bainbridge, 2011) and demonstrate the link between educational theories and evidence and classroom practice (Le Cornu & Ewing, 2008).

Our findings also emphasize the importance of additional perspectives in the analysis of knowledge applications, such as the qualities of knowledge according to de Jong and Ferguson-Hessler (1996). Analysis methods focusing on knowledge *types* did not detect changes in the learners' knowledge base in an earlier study (Klein et al., 2015).

However, even in an improved analysis, some effects did not emerge as hypothesized. The decrease in the structure of knowledge scores in the *application prob-*

lem from t2 to t3 bears closer scrutiny. Students in the integrated seminar concept were able to improve their overall task performance (with regard to the automation level of knowledge), but only to retain their t2 structure scores. Still, with regard to the knowledge taxonomy of de Jong and Ferguson-Hessler (1992), the learners' knowledge had a higher *degree of compilation* (see section 2.4) at the end of SP2. The regular seminar group's results show that this change was not only due to testing- or carry-over-effects from the previous tests. With regard to task performance, the regular seminar group's results did not improve. While they, just like the experimental group, took less time to work on the post-test application problem, their post-test structure scores were inferior to their pre-test results, i.e., their knowledge base remained at a lower degree of compilation.

The results for the transfer test were more in line with our hypotheses, with an advantage of the integrated seminar concept as well as a significant effect of the instructional measure on the structure and elaboration level of knowledge. The missing interaction, however, seen in conjunction with the instructionally supported group's performance in both seminar concepts, suggests that the effect of the instructional support measure outweighed the effects of the seminar concept and even compensated for its absence in the regular seminar group with regard to the elaboration level of knowledge. This shows that the construction of the instructional support measure along the requirements of effective instructional support during testing as outlined in section 2.4. indeed led to the desired results. With regard to structure, actively working with errors in the construction of an erroneous explanation as a part of the integrated seminar concept had, at least descriptively, additional positive effects on the learner's ability to avoid errors. These results support Oser's (2007) notion of the protective properties of negative knowledge and highlight the importance of the active construction of negative knowledge (Spychiger, 2004).

6.2 Limitations

Theory-based reflection only represents a small part of teaching competences, since many facets of teaching can hardly be taught adequately at universities (Neuweg, 2007). However, we consider it an important foundation for successful classroom acting, since it can at the very least provide a different perspective and a more substantiated knowledge base for interpretation than an entry-level teacher's own experiences. Even experienced teachers may profit from a re-evaluation of their strategies and practices in the light of scientific insights into their profession (ibid.).

The interpretation of the learning outcomes requires a reconsideration of the design principles that were used to construct the enhanced seminar, in particular, a comparison of the theoretical requirements or guidelines regarding their operationalization with ours. Didactically, the effects of *blended learning* and *fading of instructional* support in the integrated seminar conception may have canceled each other out to some degree, preventing the integrated seminar from reaching its full effectiveness. An integral part of blended learning, the self-regulated learning phases, requires learners to be instructionally supported by comprehensive work instructions and access to external resources (such as feedback from teachers, supported transfer phases; cf. Mandl & Kopp, 2006). These resources may have been reduced too quickly by the fading process in the learning phase. As is, the fading process consisted of a mere two levels instead of a more gradual, adaptive reduction (Salden et al., 2010), which may have left participants in need of additional assistance struggling.

From a methodological perspective, additional test scenarios should be constructed to improve the reliability of the measurements and minimize motivational problems. Relying on identical tests (for the application problem) allows for a comparison of the repeated measurements, however, the participants might not have been sufficiently motivated to cope with the same problem twice. Even allowing for the fact that some students scored close to the theoretical maximum at t2 and thus were unlikely to improve their explanations substantially at t3, the overall means of both groups as well as the marked decrease in the regular seminar group point toward a potential motivational problem due to the identical test scenarios at both times of measurement.

Also because of the test design, the effect of instructional support during testing on the *automation level* of knowledge could not be assessed. From an experimental design perspective, the participants' referral to the glossary, even if it was didactically unobjectionable, may have distorted the time-on-test data and thus confounded any results on the automation level for the transfer test. Therefore, we decided to focus on the effect on the structure and elaboration level of knowledge.

6.3 Conclusion: Theoretical and pedagogical implications

From a methodological position, the knowledge analyses of the present study demonstrates an innovative perspective on *applicable* knowledge according to Krause (2007). The concept of *qualities* of knowledge, based on the comprehensive knowledge taxonomy by de Jong and Ferguson-Hessler (1996), allows for the detection of effects that might have remained undiscovered by a focus on knowledge *types* by adding a second level to the analysis of knowledge applications.

From a pedagogical perspective, a more incremental or adaptive implementation of fading of instructional support might yield better results (cf. Salden et al., 2010).

Finally, while the results with regard to theory-based *reflection* are encouraging, the practical relevance of the learning environment in SP1 with regard to professional acting and decision making has to be improved. A version of the learning environment that is currently under development will include the evaluation and selection of different courses of action to remedy the problems described in the scenarios (*goal attainment* and *impact assessment/prognosis*; Beck & Krapp, 2006).

References

Agarwal, P. K., & Roedinger, H. L. III. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory, 19*(8), 836–852. doi: 10.1080/09658211.2011.613840

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060.

Bainbridge, A. (2011). Beginning teaching: The theory/practice divide. *Cliopsy, 6*, 25–32.

Bauer, J., Prenzel, M., & Renkl, A. (2015). Evidenzbasierte Praxis – Im Lehrerberuf?! Einführung in den Thementeil [Evidence-based practice – In teaching?! Introduction to the topic]. *Unterrichtswissenschaft [Science of teaching], 43(3),* 188–192.

Beck, K., & Krapp, A. (2006). Wissenschaftstheoretische Grundfragen der Pädagogischen Psychologie. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie* [*Educational Psychology*] (pp. 31–71). Weinheim: Beltz

Boshuizen, H. P. A. & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153–184.

Brightwell, R., Daniel, J., & Stewart, A. (2004). Evaluation: Is an open book examination easier? *Bioscience Education e-Journal*, *3*(1). http://www.bioscience.heacademy.ac.uk/journal/vol3/Beej-3-3.pdf

Bromme, R., Prenzel, M. & Jäger, M. (Eds.). (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. Eine Analyse von Anforderungen an die Darstellung, Interpretation und Rezeption empirischer Befunde [Empirical educational research and evidence-based educational policy] [27-2014]. *Zeitschrift für Erziehungswissenschaft [Journal for educational science],* 17(4), 3–54. doi: 10.1007/s11618-014-0514-5

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

Curry, L. (2004). The effects of self-explanations of correct and incorrect solutions on algebra problem solving performance. In K. Forbus, D. Gentner, and T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society,* 1548. Mahwah, NJ: Erlbaum.

De Jong, T., & Ferguson-Hessler, M. G. (1995). *Types and qualities of knowledge. An application to physics problem solving*. IST Internal Memo, 95-05. Enschede, The Netherlands: University of Twente, Faculty of Educational Science and Technology.

De Jong, T. & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105–113.

Durkin, K. & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3)*,* 206–214.

Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors*. Retrieved from http://hfs.sagepub.com/content/53/6/749.full.pdf+html

Evidence-based Medicine Working Group (EBMWG) (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA, 268*(17), 2420–2425.

Eysink, T. H. S., de Jong, T., Berthold, K., Kollöffel, B., Opfermann, M. & Wouters, P. (2009). Learner performance in multimedia learning arrangements: an analysis across instructional approaches. *American Educational Research Journal, 46*, 1107–1149.

Gilbert, D. T. & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38.

Greeno, J. G., Collins, A. M. & Resnick, L. B (1996). Cognition and learning. In D.C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York, NY: MacMillan.

Groeben, N., Wahl, D., Schlee, J., & Scheele, B. (1988). *Das Forschungsprogramm Subjektive Theorien [The research program Subjective Theories]*. Tübingen: Francke.

Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education, 16*(8), 811–826.

Ioannidou, M. K. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation, 23*(2), 131–139.

Kerres, M. & Jechle, T. (2002). Didaktische Konzeption des Telelernens [Didactic concept of e-learning]. In L. J.Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet. Lehrbuch für Studium und Praxis [Information and Learning with multimedia and the internet. Handbook for education and practice]* (pp. 266–281). Weinheim: PVU.

Kirschner P. A., Sweller J. & Clark R. E. (2006) Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Klein, M., Wagner, K., Klopp, E. & Stark, R. (2015). Förderung anwendbaren bildungswissenschaftlichen Wissens bei Lehramtsstudierenden anhand fehlerbasierten kollaborativen Lernens: Eine Studie zur Replikation bisheriger Befunde sowie zur Nachhaltigkeit und Erweiterung der Trainingsmaßnahmen [Fostering of applicable educational knowledge in student teachers by error-based collaborative learning: A study on the replication of earlier results and extension of the training]. *Unterrichtswissenschaft [Science of teaching], 43*(3), 225–244.

KMK – Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [standing conference of the ministries of cultural affairs in the Federal Republic of Germany]. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004) [Standards of teacher education: educational science (resolution of the conference of the ministries of cultural affairs from 16.12.2004)]*. Bonn: KMK. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

König, S. & Dalbert, C. (2007). Ungewissheitstoleranz und der Umgang mit beruflich ungewissen Situationen im Lehramt. [Uncertainty tolerance and dealing with job related uncertainty in school teachers] *Empirische Pädagogik [Empirical Pedagogy], 21*(3), 306–321.

Kopp, V., Möltner, A. & Fischer, M. R. (2006): Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. [Key-feature problems to test procedural knowledge]. *GMS Zeitschrift für Medizinische Ausbildung [GMS Journal for medical education], 23*(3), Doc 50.

Krause, U.-M. (2007). *Feedback und kooperatives Lernen [Feedback and cooperative learning]*. Münster: Waxmann.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*(5), 810–824.

Le Cornu, R., & Ewing, R. (2008). Reconceptualising professional experiences in pre-service teacher education. Reconstructing the past to embrace the future. *Teaching and Teacher Education, 24*(7), 1799–1812.

Mandl, H. & Kopp, B. (2006). *Blended Learning: Forschungsfragen und Perspektiven* (Forschungsbericht Nr. 182) [*Blended Learning: research issues and perspectives (research report no. 182*]. LMU München: Lehrstuhl für Empirische Pädagogik

und Pädagogische Psychologie [department for empirical pedagogy and educational psychology], ISSN 1614–6336.
Marton, F. & Säljö, R. (1976). On qualitative differences in learning. I – Outcome and Process. *British Journal of Educational Psychology, 46*(1), 4–11.
Meier, A. (2006). Theorienutzungskompetenz als Schlüsselqualifikation von Lehrkräften [Theory application competence as a core qualification of teachers]. In Y.M. Nakamura, C. Böckelmann & D. Tröhler (Eds.). *Theorie versus Praxis*? *[Theory versus practice?]* (S. 89-106). Zürich: Pestalozzianum.
Mekala, S. (2011). Open-book examination, a paradigm shift. *Journal of literature, culture and media studies, 3*(5/6), 263–274.
Nausner, E. (2010). *Klassenführung lernen. Praxislehrer/-innen als Ressource für den Kompetenzerwerb im Schulpraktikum [Learning classroom management. Practice teachers as a resource for competence acquisition in school internships].* Dissertation. Alpen-Adria-Universität Klagenfurt.
Neuweg, G. H. (2007): Wie grau ist alle Theorie, wie grün des Lebens goldener Baum? LehrerInnenbildung im Spannungsfeld von Theorie und Praxis. [Grey, dear friend, is all theory, and green the golden tree of life. Teachers' education in the conflict between theory and practice], *bwp@*, Ausgabe 12. Retrieved from http://www.bwpat.de/ausgabe12/neuweg_bwpat12.pdf.
Ohlsson, S. (1992). The cognitive skill of theory articulation: A neglected aspect of science education. *Science & Education, 1*, 181–192.
Oser, F. (2007). Aus Fehlern lernen [Learning from errors]. In: M. Göhlich, Ch. Wulf & J. Zirfas (Eds.): *Pädagogische Theorien des Lernens [Pedagogical theories of Learning]* (pp. 203–212). Weinheim, Germany: Beltz.
Reinmann, G. & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten [Teaching and designing learning enviroments]. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie [educational psychology]* (pp. 613-658). Weinheim, Germany: Beltz.
Renkl, A. (2006). Träges Wissen [Inert knowledge]. In: Detlev H. Rost (Ed.) *Handwörterbuch Pädagogische Psychologie [Concise Dictionary of pedagogical psychology]* (pp. 778–781). Weinheim, Germany: Beltz.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1–37.
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). Cambridge, England: Cambridge University Press
Rumelhart, D. E. (1980). Schemata: the building blocks of cognition. In: R.J. Spiro et al. (Eds.) Theoretical Issues in Reading Comprehension. Hillsdale, NJ: Lawrence Erlbaum.
Salden, R. J. C. M., Aleven, V., Schwonke, R. & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*(3), 289–307.
Schmidt, H. G, Loyens, S. M. M, Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark. *Educational Psychologist, 42*(2), 91–97.
Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.
Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5–14.
Stark, R. (2001). *Analyse und Förderung beispielbasierten Lernens: Anwendung eines integrativen Forschungsparadigmas [Analysis and facilitation of example-based learning: Application of an integrated research paradigm].* Unveröffentlichte Habilitationsschrift [unreleased postdoctoral thesis], Ludwig-Maximilians-Universität München.

Stark, R. (2005). Constructing arguments in educational discourses. In H. Gruber, C. Harteis, R. Mulder & M. Rehrl (Eds.), *Bridging individual, organisational, and cultural aspects of professional learning* (pp. 64-71). Regensburg, Germany: S. Roderer.

Stark, R., Herzmann, P. & Krause, U.-M. (2010). Effekte integrierter Lernumgebungen – Vergleich problembasierter und instruktionsorientierter Seminarkonzeptionen in der Lehrerbildung [effects of integrated learning environemnts – a comparison of problem-based and instuctional seminar designs in teacher education]. *Zeitschrift für Pädagogik [Journal of Pedagogy], 56*(4), 548–563.

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.

Steffens, D. & Reiß, M. (2009): Blended Learning in der Hochschullehre. Vom Nebeneinander der Präsenzlehre und des E-Learning zum integrierten Blended Learning-Konzept [Blended Learning in university education. From parallel existence of classroom teaching and e-larning towards an integrated Blended-Learning concept]. *HSW – das Hochschulwesen [university education], 57*(4), 115–123.

Theophilides, C., & Dionysiou, O. (1996). The major functions of the open-book examination at the university level: A factor analytic study. *Studies in Educational Evaluation, 22*(2), 157–170.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York, NY: MacMillan.

Toulmin, S. (1958). The Uses of ArgumentCambridge University Press. *Cambridge, UK*.

VanLehn, K., Siler, S., Murray, R. C., Yamauchi, T., & Baggett, W.B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249.

Wagner, K., Klein, M., Klopp, E. & Stark, R. (2014a). Theoretisieren für die Praxis: Förderung anwendbaren pädagogischen Wissens anhand advokatorischer Fehler [Theorizing in practice: Fostering of applicable pedagogical knowlegde by advocatory errors]. *Journal für Lehrerinnen- und Lehrerbildung [Journal for teacher education], 1*, 65–74.

Wagner, K., Klein, M., Klopp, E. & Stark, R. (2014b). Instruktionale Unterstützung beim Lernen aus advokatorischen Fehlern in der Lehramtsausbildung: Effekte auf die Anwendung wissenschaftlichen Wissens [Instructional support in learning from advocatory errors in teacher education: Effect on the application of scientific knowledge]. *Psychologie in Erziehung und Unterricht, 61*, 287–301.

Wecker, C. (2012). *Vom Sollen zum Können. Fading instruktionaler Skripts zur Förderung von Argumentationskompetenz [From target to ability. Fading of instructional scripts to support competence of reason]*. Münster, Germany: Waxmann.

Wecker, C. & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. *Learning and Instruction, 21*(6), 746–756.

Wenglein, S., Bauer, J., Heininger, S. & Prenzel, M. (2015). Kompetenz angehender Lehrkräfte zum Argumentieren mit Evidenz: Erhöht ein Training von Heuristiken die Argumentationsqualität? [Pre-Service Teachers' Evidence-Based Argumentation Competence: Can a Training of Heuristics Improve Argumentative Quality?] *Unterrichtswissenschaft [Science of teaching], 43*, 209–224.