Wedde, Sonja; Busse, Annette; Bosse, Dorit

Comparing contrasting cases in problem-solving first activities. Insights into analytical solution quality

Journal for educational research online 16 (2024) 2, S. 261-280

Quellenangabe/ Reference:
Wedde, Sonja; Busse, Annette; Bosse, Dorit: Comparing contrasting cases in problem-solving first activities. Insights into analytical solution quality - In: Journal for educational research online 16 (2024) 2, S. 261-280 - URN: urn:nbn:de:0111-pedocs-346976 - DOI: 10.25656/01:34697; 10.31244/jero.2024.02.05

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-346976
https://doi.org/10.25656/01:34697

in Kooperation mit / in cooperation with:

WAXMANN
www.waxmann.com

http://www.waxmann.com

Journal for Educational Research Online
Journal für Bildungsforschung Online
https://doi.org/10.31244/jero.2024.02.05
Volume 16 (2024), No. 2, S. 261–280

Sonja Wedde, Annette Busse & Dorit Bosse

Comparing Contrasting Cases in Problem-Solving First Activities: Insights Into Analytical Solution Quality

Abstract

In educational research, there are promising findings which suggest it is useful to have students work on a task before instruction takes place. This also involves using task formats that follow the principle of comparison with and without predefined categories. To date, few studies have evaluated the learning processes induced by comparing contrasting cases before instruction takes place. As comparing requires further analytic cognitive processes, such as recognizing, differentiating, and organizing, a construct based on analytic competence was developed to evaluate 149 student teachers' task solutions using qualitative content analysis. One experimental group compared cases using self-generated categories (invention activity); the other group was given categories to compare the cases (worked solution). After comparing the cases, the students were introduced to classroom management, a topic that is relevant for prospective teachers. The purpose of this task was to acquire knowledge about and develop a professional vision of classroom management. In addition, to evaluating participants' task solutions, a standardized questionnaire was used to collect data on the extrinsic and intrinsic cognitive load of the students. Students in the worked solution condition achieved significantly higher analytic solution quality and experienced significantly lower intrinsic cognitive load than students in the invention activity condition. Both experimental groups perceived an average low extrinsic cognitive load.

Dr. Sonja Wedde, Institute of Educational Science, University of Kassel,
Nora-Platiel-Str. 1, 34109 Kassel, Germany
e-mail: sonja.wedde@uni-kassel.de

Dr. Annette Busse, Institute of Educational Science, University of Kassel,
Nora-Platiel-Str. 1, 34109 Kassel, Germany
e-mail: abusse@uni-kassel.de

Prof. Dr. Dorit Bosse, Institute of Educational Science, University of Kassel,
Nora-Platiel-Str. 1, 34109 Kassel, Germany
e-mail: bosse@uni-kassel.de

Keywords
analytical solution quality, invention activities, worked solutions, cognitive load, professional vision

Vergleichen kontrastierender Fälle in Problem-Solving first Aktivitäten: Erkenntnisse zur analytischen Bearbeitungsqualität

Zusammenfassung
In der Lehr-Lernforschung gibt es vielversprechende Befunde, nach denen es sinnvoll erscheint, Lernende eine Aufgabe bearbeiten zu lassen, bevor die Instruktion erfolgt. Dabei kommen auch Aufgabenformate zum Einsatz, die dem Prinzip des Vergleichens mit und ohne vorgegebene Kategorien folgen. Bisher liegen wenige Studien zur Auswertung von Lernprozessen vor, die durch das Vergleichen kontrastierender Fälle ausgelöst werden, bevor die Instruktion erfolgt. Da Vergleichen analytischer kognitiver Prozesse wie erkennen, differenzieren und organisieren bedarf, wurde ein Konstrukt basierend auf der Analysekompetenz entwickelt, um 149 Aufgabenbearbeitungen von Lehramtsstudierenden mittels qualitativer Inhaltsanalyse auszuwerten. Eine Experimentalgruppe verglich die Fälle mittels selbstentwickelter Kategorien (Invention Activity), die andere Gruppe erhielt Kategorien, um die Fälle zu vergleichen (Worked Solution). Nachdem die Studierenden die Fälle verglichen hatten, erhielten sie eine Einführung in das für angehende Lehrpersonen bedeutsame Thema Klassenführung. Ziel der Aufgabe war es Wissen zu Klassenführung zu erwerben sowie die professionelle Wahrnehmung von Klassenführung anzubahnen. Zusätzlich zur Auswertung der Aufgaben wurden mithilfe eines standardisierten Fragebogens Daten zur extrinsischen und intrinsischen kognitiven Belastung der Studierenden erhoben. Die Studierenden in der Bedingung Worked Solution erreichten eine signifikant höhere analytische Bearbeitungsqualität und empfanden eine signifikant geringere intrinsische kognitive Belastung als die Studierenden in der Bedingung Invention Activity. Beide Experimentalgruppen nahmen eine durchschnittlich geringe extrinsische kognitive Belastung wahr.

Schlagworte
Analytische Bearbeitungsqualität, Invention Activities, Worked Solutions, kognitive Belastung, professionelle Wahrnehmung

1. Introduction

Research on teacher education has studied the effectiveness of different task formats in initiating meaningful learning processes for student teachers. Two promising task formats are invention activities and worked solutions at the beginning of a new learning unit, which, however, have been little researched. An invention activity is a learning method in which learners are asked to generate a solution to a given task by working with contrasting cases. Subsequently, the instruction phase involves presenting the canonical solution and instruction on the new topic (Schwartz et al., 2011). This task format is based on the problem-solving prior to instruction approach and is considered promising for raising epistemic curiosity, activating prior knowledge, and enabling awareness of knowledge gaps at the beginning of a learning unit (Loibl et al., 2017). Worked solutions involve learners being presented with a sample solution to the contrasting cases, which they must study whilst working on the task (Renkl, 2014).

To date, research on invention activities and worked solutions has mainly focused on learning outcomes, the quality of the solution, and aspects such as epistemic curiosity or extraneous cognitive load. Different aspects of solution quality were evaluated in those studies. To compare contrasting cases, learners have to recognize relevant features, arrange them accordingly, and relate them to each other in the given task. These cognitive processes have not been analyzed in studies on solution quality regarding invention activities. The focus of this study is the analysis of these cognitive processes by using an in-depth approach called analytical solution quality. This construct is based on the analytical competence since the cognitive processes when comparing contrasting cases are based on analytical processes (Plöger et al., 2020).

The task at hand contains two audio classroom examples focusing on classroom management. Young teachers often experience difficulties with effective classroom management in their teaching practice (Chaplain, 2008). Therefore, this task aims to promote knowledge about classroom management as well as a professional vision of classroom management. Professional vision is a situation-specific cognitive skill that mediates between dispositions, professional knowledge and affect motivation, and classroom performance (Blömeke et al., 2015; Blömeke & Kaiser, 2017). Thus, professional vision should already be promoted in university as it functions as a mediator between professional knowledge and classroom performance.

Until now, no research on teacher education has examined whether comparing a successful teaching example to a less successful teaching example is more beneficial when the comparison is based on a set of provided criteria (worked solution) or self-generated criteria (invention activity). Therefore, from the framework of analytical competence (Plöger et al., 2020), we derived a model to analyze the teacher students' comparison processes. By coding student teachers' solutions, more insight will be gained into the comparative processes of contrasting cases of learners. Data from an introductory lecture in educational studies were used to gain insights on which learning format is more effective. Additionally, we combined the coded data with data on the extraneous and intrinsic cognitive load.

2. Invention Activities and Worked Solutions

Although the meta-analyses by Alfieri et al. (2013), Apthorp (2010), and Marzano et al. (2001) identified that compare-and-contrast activities have a high learning impact on learning at school and university, other formats that integrate comparison activities into invention activities have been shown to yield better learning outcomes than the simple identification of similarities and differences (Chi et al., 2012; Chin et al., 2016). Overall, invention activities promote the acquisition and transfer of conceptual knowledge in particular (Loibl et al., 2017).

In comparisons between invention activities and other formats, the cognitive load theory is often considered. According to some scholars, the problem-solving prior to instruction approach induces excessive cognitive load, especially amongst beginners, which hinders learning processes (Kirschner et al., 2006). Therefore, worked solutions are often chosen as the second experimental condition for comparison with invention activities, as working with worked solutions presumably results in lower cognitive load. Cognitive load is divided into extraneous and intrinsic cognitive load. Extraneous cognitive load does not contribute to learning and refers to the cognitive load associated with the design of the task and its materials (Sweller et al., 1998). A study involving eighth graders showed that the group who completed an invention activity perceived significantly more extraneous cognitive load than the group who worked on a worked solution (Glogger-Frey et al., 2015). In contrast, our previous study found no differences in perceptions of extraneous cognitive load between the two experimental groups (Wedde et al., 2021).

Besides extraneous cognitive load, intrinsic cognitive load, which is dependent on learners' prior knowledge and element interactivity, is also used to assess the effectiveness of a learning process (Sweller et al., 1998). In general, element interactivity refers to the complexity of the learning material. More specifically, it refers to elements that learners must simultaneously process in their working memory to complete a task; for example, in mathematics, these elements include the number of arithmetic operations that need to be kept in working memory to solve an equation (Ashman et al., 2020).

Several studies in problem-solving prior to instruction research have reported inconsistent results (e.g. Glogger-Frey et al., 2017; Loibl et al., 2020). Other studies, which did not compare the invention activity with a worked solution condition, attributed positive learning outcomes to invention activities (e.g. Chi et al., 2012; Chin et al., 2016; Schwartz et al., 2011). However, research has demonstrated that, in comparison to invention activities, worked solutions may lead to better learning outcomes for both student teachers and students (Glogger-Frey et al., 2015; Glogger-Frey et al., 2022). In their study of tenth- and eleventh-grade students, Chen et al. (2016) reported that problem-solving led to better learning outcomes when element interactivity was low and worked solutions led to better learning outcomes when element interactivity was high. High element interactivity means that learners would be cognitively overloaded and effective learning would be hindered. However, when element interactivity was low, formats based on the problem-solving prior to instruction approach were more suitable (Ashman et al., 2020). When element interactivity was high, studies favored worked solutions (Chen et al., 2016; Glogger-Frey et al., 2015). Therefore, Ashman et al. (2020) recommended placing direct instruction before problem-solving formats when element interactivity was high. Overall, results on the effectiveness of invention activities have been inconsistent. Intrinsic cognitive load, and thus also the level of element interactivity, may be one explanation for differences in study results. Thus far, most studies have been conducted in science and mathematics and generally in school contexts.

Different aspects of solution quality have been evaluated in several studies. For example, Loibl and Rummel (2014) verified the agreement between the number of items in task solutions with those of a canonical solution. The experimental conditions in their study differed from those in our project. One group was tasked with developing a solution and was equipped with contrasting cases as guidance whilst the other group had to invent solutions without guidance in the form of contrasting cases. The researchers found the group that received guidance in the form of contrasting cases generated better solutions than the one that did not receive extra support. Moreover, Glogger-Frey et al. (2015) assessed the appropriateness of task solutions only for invention activities, not for worked solutions. In addition, Wiedmann et al. (2012) categorized and assessed answers according to the mathematical concept to be learned. Their study focused on group compositions for working on the tasks. Therefore, it did not include a worked solution experimental condition. One finding from their study was that high-quality solution attempts were related to the outcome in the posttest.

3. Modelling Analytical Solution Quality

Since comparing cases contains analytical and classification activities (Hilker, 1962), the construct of analytical solution quality used in this study was based on Plöger et al.'s (2020) model of analytical competence and related considerations. Comparatively evaluating auditive teaching examples requires analytical competence. It was, thus, assumed that analytical and comparative processes cannot be regarded as distinct from each other; rather, comparative processes are based on analytical processes. Hence, as a task format, comparison can also be understood as training to improve the analytical competence of prospective teachers.

According to Plöger et al. (2020), analytical competence is defined as the ability to perceive and evaluate the quality and learning effectiveness of observed classroom teaching. The analytical competence consists of two dimensions: the content dimension, which consists of the knowledge facets of pedagogical knowledge and content knowledge, and the formal dimension, which the researchers called *complexity of information processing*. In the present study, we only refer to the formal dimension to assess the analytical quality of students' solutions. The stages of the formal dimension reflect an increasing degree of complexity in information processing. Plöger et al.'s (2020) study yielded consistent results: experts achieved higher competence in the analysis of lessons than novices.

Based on the expertise paradigm (Berliner, 2004), the following considerations were included in the model of analytical competence and are essential to the construct of analytical solution quality. First, experts are more likely to notice deep features of teaching, such as student learning, the interaction between teacher and students, and classroom management strategies. Novices are more likely to observe surface features such as the media used in class or the form of social interaction (Berliner, 2004; Wolff et al., 2016). It is assumed that surface features – in contrast to deep features – are directly observable. Studies have shown that learning is more related to deep features than to the surface features in teaching (Kunter et al., 2013). Comparing successful and less successful examples potentially enables student teachers to understand that teaching effectiveness depends on deep features rather than surface features, provided that aspects of these deep features are varied in the examples. For classroom management, deep features refer, for example, to the management of disruptions, the linking of rules and instruction, or the organization of transitions. Surface features refer to more obvious aspects of classroom management, such as the existence of rules and routines (Wedde et al., 2022).

Second, experts and novices differ in their observation processing (Plöger et al., 2020). Experts can recognize the complexity of teaching and thus categorize individual actions and situations as part of a larger context. This ability is due, in particular, to their highly developed knowledge (including case knowledge). Novices, on the other hand, tend to perceive lessons in parts and thus view situations and actions independently of the lesson's overall context (Carter et al., 1988; König et al., 2014). Based on these considerations, the levels of analytical solution quality are outlined below.

Comparing contrasting cases involves more cognitive processes than simply "comparing" (e.g. recognizing, inferring, differentiating, organizing, generating; Anderson, 2014). It can be understood as an activity that encompasses observational, analytical, and classifying tasks and can be described as relational thinking (Hilker, 1962). Thus, a reference ratio is established between the objects to be compared. There can be sameness, resemblance, or dissimilarity between them. Table 1 summarizes different comparison models including Hilker's (1962) and Bereday's (1964)[2] four steps of comparison for comparative education, Wilcke and Budke's (2019) model for geography education, and also the model developed in this study as described below. Hilker's (1962), Bereday's (1964), and Wilcke and Budke's

2 Adick (2018) indicated that Hilker and Bereday's models of comparison were developed in parallel but included the same steps.

(2019) models were used to develop the construct of analytical solution quality (see Table 1). This construct aims at gaining a better understanding of the comparison processes of contrasting cases in invention activities.

Table 1: *Steps of Comparison in Different Models*

Step	Hilker (1962) and Bereday (1964) Bereday Comparative education	Wilcke and Budke (2019) Geography education	The model presented in this study Educational studies
1	Description	Developing a question	Description
2	Interpretation	Determining units of comparison	Classification into categories
3	Juxtaposition	Definition of variables for comparison	Juxtaposing
4	Comparison	Juxtaposing similarities and differences	Summarization
5		Weighing comparison variables, developing connections between units of comparison and deriving explanations	Conclusion
6		Evaluating the results and answering the question	

The construct of analytical solution quality comprises five levels that represent the comparative depth of task solutions: (1) *description*, (2) *classification into categories*, (3) *juxtaposition*, (4) *summarization* and (5) *conclusion* (see Table 1). The steps follow the assumptions described in Commons et al.'s (1998) model of hierarchical complexity. To reach a higher step, the preceding steps must be executed. Thus, the execution of the higher steps results from connecting lower steps. The individual steps are interrelated in a non-arbitrary way and build on each other. This interrelationship increases the degree of complexity of information processing across the individual steps.

In this section, the steps of the comparison process are described for classroom management, the subject chosen for this project. Similar to Hilker (1962) and Bereday (1964), we also consider *description* to be the first step in comparison. *Description* means that students can refer to individual situations or actions that are directly related to classroom management. These descriptions can be actively included in the comparison. This step refers to the surface structure of teaching, meaning that the learners first describe on a surface level what they have perceived.

The second step, *classification into categories*, concerns the deep structure of teaching. In this step, perceived aspects that have already been described are classified into concepts related to student learning. In other models of comparison, the aspects to be compared are also classified into categories. Wilcke and Budke (2019) described this process in the step of juxtaposing similarities and differences, whilst Hilker (1962) used a similar description of juxtaposition for comparative education and regarded it as the first stage of actual comparison. In our construct, this step involves students matching situations and actions to categories. In this step, they can, thus, assign situations and actions to techniques or features of classroom management and organize and process the information at a higher level.

The third step, *juxtaposing*, means that different situations and actions from the examples are compared related to each other and that their effects and consequences are appropriately assessed and evaluated. This step transcends mere classification into categories. It is desirable for students to consider the different perspectives of and the dynamic between teacher and students.

Summarization is similar to the fifth step in Wilcke and Budke's (2019) model, in which connections must be made between units of comparison (i.e. the material to be compared). In this step, for our research, a brief summary is drawn about which of the two auditive lesson examples represented a better use of classroom management strategies. The step also includes a final evaluation of perceived actions and classroom management situations.

The final step in Hilker (1962) and Bereday's (1964) model, comparison, encompasses the fourth and fifth steps in our construct and involves summarizing and generalizing results. In Wilcke and Budke's (2019) model, the last step also involves the evaluation of results, which is similar to the fourth step in our construct, *summarization*.

Conclusion is based on and follows the evaluation in the *summarization*. It entails assessing and evaluating individual actions and situations, particularly with regard to learning effectiveness. Thus, actions were not considered in isolation but rather alongside other actions as part of a larger context. Finally, effective classroom management strategies were to be identified for one's own teaching practice. The fifth step, *conclusion*, is particularly important for practical relevance in teacher education. Student teachers presumably learned effective techniques for their own teaching practice from the presented examples.

The models of comparison operate at different levels. Hilker (1962) and Bereday's (1964) model aims to yield scientific knowledge, whilst Wilcke and Budke's (2019) model focuses on increasing students' competence and expanding students' geographical knowledge. By contrast, our construct aims to promote analytical competence amongst students (and, by extension, a professional vision of classroom management) and introduce effective classroom management techniques through the comparison of two practical classroom examples.

Young teachers often face difficulties in applying effective classroom management strategies (Chaplain, 2008). Therefore, promoting a professional vision of classroom management could enable prospective teachers to recognize actions that benefited or hindered learning in the classroom at an early stage (*noticing*) and to interpret them in a theory-based manner (*knowledge-based reasoning*) to react appropriately to the situation (Gold et al., 2020; Sherin, 2007). These two interrelated, knowledge-based sub-processes are part of teachers' professional competence (Sherin, 2007) and considered to be trainable (Stürmer et al., 2013). To notice and interpret relevant events in the classroom, teachers need declarative, conceptual, and case knowledge (Berliner, 2001; König et al., 2014; Stürmer et al., 2013). By working on the task within the problem-solving prior to instruction approach, students first gain case knowledge by comparing the two auditive classroom examples. During the subsequent instruction, declarative as well as conceptual knowledge about classroom management is promoted. Thus, a professional vision of classroom management, and thus also the analytical competence, is developed. There is some evidence to show that comparison positively affects how professional competence is acquired (Heemsoth & Kleickmann, 2018): Comparing effective and less effective teaching examples improves lesson planning competence in the area of physical education and supports the development of constructivist beliefs.

There are several models of professional vision with different facets of competence. Comparing the steps in our construct with facets of competence from different models of professional vision allows common features to be identified. In Sherin and van Es's (2009), Seidel and Stürmer's (2014), and Gold et al.'s (2016) models, the first step is *describing*. Other aspects of these models include *evaluating* and *interpreting* (Sherin & van Es, 2009), and *explaining* (Seidel & Stürmer, 2014). These aspects are also found in the steps of analytical solution quality. Therefore, it is assumed that comparing auditive teaching examples can promote a professional vision amongst students.

4. Research Questions

The objectives of the present study were two-fold. First, it aimed to determine the analytical quality of solutions developed by students in the two experimental conditions, invention activities and worked solutions. Second, the study aimed to determine differences in perceptions of extraneous and intrinsic cognitive load between participants in both experimental conditions and the relationship between intrinsic cognitive load and analytical solution quality. To this end, we formulated the following research questions:

1. How do the experimental conditions, invention activities, and worked solutions differ in terms of analytical solution quality?
2. How do the experimental conditions, invention activities, and worked solutions differ in terms of extraneous and intrinsic cognitive load?
3. Is there a correlation between intrinsic cognitive load and analytical solution quality?

For the first research question, we assumed that there would be differences between the experimental conditions: worked solutions (WS) and invention activities (IA). The WS group received more support during the task than the IA group because its participants were provided with a set of predefined categories. Thus, the following hypothesis was developed: Participants in the WS group will achieve a higher analytical solution quality than those in the IA group (H1).

Although we did not find a difference regarding extraneous cognitive load between the two experimental groups in a previous study (Wedde et al., 2021), from a theoretical point of view, there might be a difference between both experimental groups, as also confirmed by Glogger-Frey et al. (2015). Therefore, we assume that students of the IA group perceive higher extraneous cognitive load than students in the WS group (H2a). Due to the large number of elements (i. e. the dimensions of classroom management, which were initially unknown to participants) and simultaneous processes that students had to contend with during the task, we assumed that there would be a high level of element interactivity, especially for the IA group, where participants had to generate their own categories. Therefore, we assumed that there would be a higher intrinsic cognitive load for the IA group than for the WS group (H2b). This difference could have had an impact on possible comparison processes during the task and thus on the quality of the analytical solution. We, therefore, formulated the following hypothesis for the third research question: There is a negative correlation between intrinsic cognitive load and analytical solution quality for the overall sample, such that the higher the intrinsic cognitive load, the lower the analytical solution quality (H3).

5. Method

5.1 Sample

The sample consisted of 149 student teachers (65.8% female) in the introductory phase of their studies at the University of Kassel, Germany. Only cases for which analytical solution quality and the cognitive load scale were available were selected from the total sample ($N = 256$). Therefore, 53 cases (41.1%) were removed from experimental group IA and 54 cases (42.5%) were removed from experimental group WS. Thus, a similar number of cases were removed from both groups, $\chi 2\ (1) = 0.54$, $p = .82$. The participants were randomly assigned in nearly equal proportions to the two experimental conditions: 76 students were assigned to the IA group (67.1% female, age: $M = 22.3$, $SD = 4.85$), whilst 73 students were assigned to the WS group (64.4% female, age: $M = 21.4$, $SD = 4.80$).

5.2 Research Design

The current experimental study was conducted in the context of an introductory course in educational studies of the teacher training program during the 2020–2021 winter term.[3] Participants in both groups were tasked with comparing two audi-

3 Due to the COVID-19 pandemic, the course was held online via video conferencing.

tive teaching examples. Afterwards, process variables' data were collected. Overall, working on the task and completing the questionnaires took 90 minutes, equivalent to one session (see Figure 1 for an overview of the data collection process).

Figure 1: *Two Experimental Conditions in the Problem-Solving Prior to Instruction Approach*

	Invention Activity $N = 76$	Worked Solution $N = 73$
Phase I	Comparing auditive classroom examples with *self-generated* categories	Comparing auditive classroom examples with *given* categories
	Data collection of process variables and demographic data, including extraneous and intrinsic cognitive load	
Phase II	Instruction and canonical solution	

5.3 Treatment

In this study, two auditive lesson examples were used as contrasting cases. Students were asked to listen to segments of constructed teaching examples. The second auditive classroom scene presented successful use of classroom management strategies by a teacher, whilst the first auditive classroom scene presented a less successful use of such strategies. Before listening to the two classroom scenes, the students could gain an overview of the task. They could listen to the two cases as often as they liked. All students had 60 minutes to complete the assignment. They wrote down the results of their comparison. The task differed between the two experimental conditions: After listening to the two auditive classroom examples, students in the IA group were asked to generate categories to classify the teacher's level of classroom management and then compare the examples according to these categories. There was no minimum number of categories that students had to generate. By contrast, students in the WS group were provided with categories (i.e. managing transitions, rules, routines, communication by the teacher, and managing disruptions) and asked to compare the two auditive classroom examples according to those pre-defined categories. The two classroom examples provided for the WS group differed in all the given categories. In another study, we evaluated how many categories the two experimental groups each compared on average. The experimental group WS mentioned more categories for both auditive classroom examples than the experimental group IA (Wedde et al., 2022; see Figure 2).

It can be assumed that working on the task required a high level of element interactivity, as students had to rely on their listening skills to process the classroom examples. In addition, the topic of classroom management was addressed, of which they presumably had little prior knowledge as first-year students. Thus, they first had to recognize situations that were relevant to classroom management, then classify them into categories. Therefore, although students in both groups were expected to process simultaneously several elements, a higher element interactivity was assumed for the IA experimental condition due to the challenge of the students having to develop their own categories.

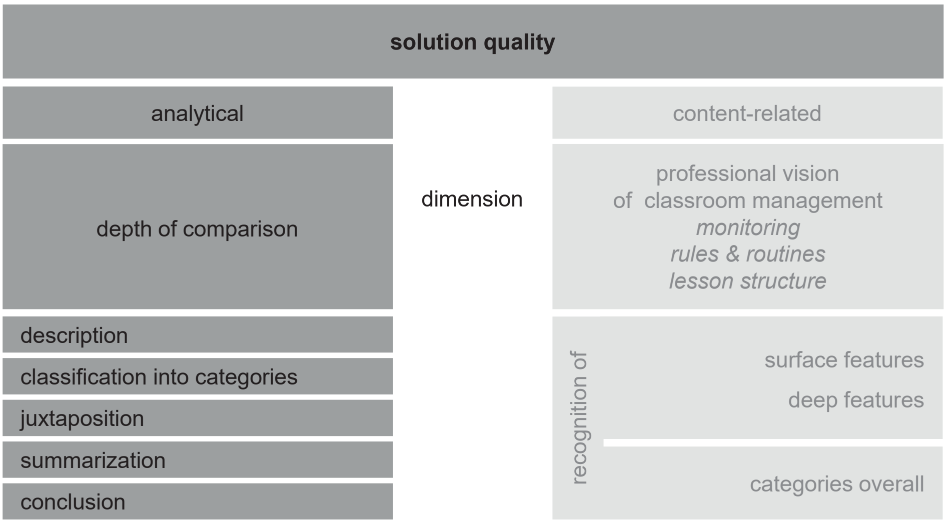
In the professional development of student teachers, contrasting cases are a method of illustrating the link between theory and practice. For classroom management, it can be useful to present contrasting cases to make less successful strategies salient compared to pedagogical valuable strategies. Since student teachers have spent many hours in classrooms during their own time in schools, they have experienced many different applications of classroom management strategies. By analyzing contrasting cases, they become aware of pedagogically valuable and effective strategies when strategies of an experienced teacher are contrasted to strategies of an inexperienced teacher (Wedde et al., 2022).

5.4 Instruments

5.4.1 Depth of Comparison Scale

The analytical analysis was performed using the construct of solution quality. Within this, we differentiated between analytical and content-related solution quality (see Figure 2; Wedde et al., 2022). For the purposes of this study, we focused on analytical solution quality. Analytical solution quality serves as an indicator of the quality of the learning processes induced by the task of comparing contrasting cases. Therefore, the focus is on capturing the learning processes induced by the task rather than on the competence that may be acquired through the task.

Figure 2: *Construct of Solution Quality*



On the basis of the levels of analytical solution quality, five categories were derived to be coded for in the students' solutions. Those categories indicate the depth of the analytical level of the solutions (see Table 2). These five categories were defined and described in a coding manual. Additionally, for each of the categories, negative indicators and anchor examples were given. Based on this coding manual, two trained coders performed the qualitative content analysis. Qualitative content analysis is characterized by a systematic process that aims to assess and evaluate cases to be analyzed based on selected categories (Mayring, 2015).

Table 2: *Categories of the Coding Manual*

Step		Condition	Definition
1	Description		Reproduction of classroom situations and/or actions relevant to classroom management
2	Classification into categories	IA 2a	References to classroom situations and/or actions from both examples based on self-generated classroom management categories for comparison
		WS 2b	Collection of specific and relevant classroom management actions and/or situations from both examples according to the given categories
3	Juxtaposition		Comparison of both examples using the self-generated or given categories
4	Summarization		Development of summaries based on the comparative analysis of which of the two examples represented a better use of classroom management strategies
5	Conclusion[a]		Development of conclusions on the comparative analysis of learning effectiveness in both teaching examples and benefits to students' future teaching practice

Note. Step 2 is divided into two steps, each relating to the two experimental conditions.
[a]Conclusion was not included in the scale as this step was not found in the students' solutions.

The categories were individually coded for the students' entire texts in a semi-holistic manner (Schipolowski & Böhme, 2016). The two coders performed the analysis of the 149 solutions separately. Due to the heterogeneity of the task solutions, there were some discrepancies between the codings. To ensure the reliability of the approach, a value was agreed upon for non-matching values with subjective assessment and, if necessary, the category definition was adjusted (Guest et al., 2011).

In a second step, all coded categories were transformed into numbers, which were derived from the coding of individual steps. Only if a step was achieved was one point given. Resultingly, a solution could be assigned to several steps. In line with the assumptions outlined above, it was necessary for a lower step to be achieved before achieving a higher step, and a total of five points could be assigned for the five steps. For example, to achieve step 3, it was also necessary to achieve steps 1 and 2, so three points were given. On this basis, the scale "depth of comparison" was created. The higher the value on the scale, the deeper the comparison. Thus, the scale was ordinal in nature. Based on the given points, a sum score was created (Plöger et al., 2020), and the scale was evaluated as an interval scale ($\alpha = .73$).

5.4.2 Cognitive Load

After completing the task, participants were asked to rate extraneous and intrinsic cognitive load on a scale from 1 ("strongly disagree") to 6 ("strongly agree"). Four items were used to measure extraneous cognitive load ($\alpha = .68$, e.g. "The task was very unclear."). Intrinsic cognitive load was measured with three items ($\alpha = .86$, e.g. "The topics covered in the podcasts were very complex."). The scales were based on Leppink et al.'s (2013) instrument and adapted to the present study. Leppink et al.'s (2013) instrument was based on Sweller et al.'s (1998) definition of extraneous and intrinsic cognitive load.

5.4.3 Professional Vision of Classroom Management

To ensure comparability of the two experimental groups, we assessed the professional vision of classroom management three weeks before treatment. For this purpose, we used a standardized video-based online test. This variable indicates a coefficient of agreement with an expert rating from 0 to 1 (Gold & Holodynski, 2017).

5.5 Data Analysis

For the instruments, mean values and standard deviations were calculated. To examine differences between the two groups, *t*-tests for independent samples were conducted. Experimental condition (IA or WS) was the independent variable, and depth of comparison, extraneous or intrinsic cognitive load was the dependent variable. In addition, a Pearson correlation coefficient was calculated for the correlation between depth of comparison and intrinsic cognitive load. A significance level of .05 was used for all analyses.

6. Results

Neither the experimental group IA ($M = .34$, $SD = .19$) nor WS ($M = .38$, $SD = .18$) differed significantly at pretest in their professional vision of classroom management, $t(144) = -1.32$, $p = .19$.

In accordance with Hypothesis 1, there was a highly significant difference in depth of comparison between the IA and WS experimental conditions, which was associated with a large effect size, $t(146.938) = 4.60$, $p < .001$, $d = 0.75$. Depth of comparison was higher for the WS experimental condition than for the IA experimental condition (see Table 4), which confirmed Hypothesis 1. These differences were also evident from the individual steps on the depth of comparison scale (see Table 3). Indeed, the differences were noticeable as early as the first step, *description*. In addition, slightly more than half of participants' task solutions reached the second step, *classification into categories*, whilst nearly a third reached the third step, *juxtaposition*. The fourth step, *summarization*, was achieved by less than 10% of participants. Overall, more participants in the WS group than in the IA group reached the higher steps of the scale.

Table 3: *Descriptive Statistics of the Individual Coded Categories*

	WS		IA		Overall	
Depth of comparison	*N*	%	*N*	%	*N*	%
Description (1)	62	84.9	44	57.9	106	71.1
Classifying into categories (2)	57	78.1	30	39.5	87	58.4
Juxtaposition (3)	28	38.4	17	22.4	45	30.2
Summarization (4)	9	12.3	3	3.9	12	8.1
Conclusion (5)	0		0		0	

Note. The lower steps of the scale must be reached before higher steps. Step 5 was not found in the solutions.

The results from extraneous cognitive load investigations show that students of both experimental groups perceived similar extraneous cognitive load (see Table 4). Thus, there was no significant difference between the two experimental groups, $t(137) = -0.19$, $p = .85$. Consequently, Hypothesis 2a must be rejected. There was a significant difference in intrinsic cognitive load between both experimental groups, which was associated with a small to medium effect size, $t(143) = -2.15$, $p = .03$, $d = 0.35$. Hence, Hypothesis 2b was confirmed. Although participants in the WS experimental condition reported a lower intrinsic cognitive load on average than those in the IA experimental condition (see Table 4), all participants scored on the lower range of the intrinsic cognitive load scale.

Even though the groups showed significant differences in both intrinsic cognitive load and depth of comparison, the correlation between these constructs was not significant, $r = -.04$, $p = .34$, $N = 145$, one-tailed. Additionally, the correlations for the

Table 4: *Means and Standard Deviations for Depth of Comparison and Cognitive Load*

Variable	WS	IA	Overall
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Depth of comparison	2.14 (1.16)	1.24 (1.23)	1.68 (1.27)
Extraneous cognitive load	1.65 (0.72)	1.63 (0.64)	1.64 (0.68)
Intrinsic cognitive load	2.09 (0.87)	2.40 (0.95)	2.25 (0.93)

Note. Items on the extraneous and intrinsic cognitive load scale are rated from 1 ("does not apply at all") to 6 ("applies completely"), whilst sum scores on the depth of comparison scale range from 0 to 4.

WS group, $r = -.03$, $p = .40$, $N = 69$, one-tailed, as well as for the IA group, $r = .08$, $p = .25$, $N = 76$, one-tailed, were not significant. Therefore, Hypothesis 3 was rejected.

7. Discussion and Conclusion

In this study, a construct was presented to determine analytical solution quality for IA and WS in a task that involved comparing two auditive classroom examples. This construct aimed at evaluating learning processes induced by comparing contrasting cases. With regard to the first research question, the results demonstrated that the WS group achieved a significantly higher analytical solution quality than the IA group. However, participants in the overall sample tended to reach only the lower range of analytical solution quality. These findings align with those of Plöger et al. (2020), in which novices demonstrated a lower level of analytical competence than experts. In addition, these results also align with research on expert teachers: Novices tend to examine situations independent of context, whereas experts tend to observe and classify lessons more comprehensively (Carter et al., 1988).

For the second research question, participants in the WS group perceived significantly less intrinsic cognitive load during the task than those in the IA group. This result was in line with our expectations; higher element interactivity was assumed for IA since no categories were provided for comparison and students had to classify the situations according to their own assessments and ideas of classroom management. Due to this classification task and their low prior knowledge of classroom management, they had to process simultaneously more elements than the WS group.

No significant difference was found between the two groups regarding extraneous cognitive load. In one of our previous studies, we came to the same finding, although a different study found that the invention activity tended to result in higher extraneous cognitive load than the worked solution (Glogger-Frey et al., 2015). Extraneous cognitive load was low for both experimental groups, indicating that, despite its complexity, the task itself did not contribute to irritation or overload among students.

However, contrary to our expectation that high element interactivity would result in high intrinsic cognitive load (Ashman et al., 2020), overall intrinsic cognitive load was low for the entire sample. Possible explanations for this finding include the methods used in the data collection or the time at which participants' intrinsic cognitive load was surveyed. In future studies, cognitive load could be assessed using fewer abstract items, and it needs to be assessed during working through the task and not after finishing the task. Items should be more specific regarding the task, e.g. in terms of inventing categories or working with given categories.

For the third research question, the correlation between intrinsic cognitive load and analytical solution quality was not significant. Since so many of the students never reached step 2, *classification into categories*, and finished at the *description* stage, it is questionable whether we measured intrinsic cognitive load with our scale. This would mean that some of the students were not engaging in the learning process nor relating items to each other, which could have resulted in intrinsic cognitive load. It would also explain why intrinsic cognitive load and analytical solution quality are not significantly related.

Nevertheless, it remains uncertain whether high analytical solution quality is associated with better learning outcomes. A future study will examine whether IA and WS are effective task formats for comparing two contrasting auditive classroom examples to promote a professional vision of classroom management. Other studies that examined the problem-solving prior to instruction approach reported mixed results on the relationship between solution quality and learning outcomes (Loibl & Rummel, 2014; Wiedmann et al., 2012).

Since Plöger et al. (2020) indicated that analytical competence is a two-dimensional construct, a formal dimension and a content-related dimension, we also coded for the content-related dimension of solution quality. Results from a previous study showed that the student teachers' solutions contained overall low content-related solution quality (Wedde et al., 2022). The results of the present study were similar. Both our studies have shown that the WS group demonstrated slightly higher quality than the IA group.

It can be considered a limitation of this study that no knowledge test was used to elicit prior knowledge about classroom management. We only assumed that the first-year teacher students had a low prior knowledge of classroom management. In future studies, prior knowledge about classroom management needs to be assessed.

Since the results from this study demonstrated that neither the IA nor the WS group achieved a particularly high level of analytical solution quality, the task should include support for future students by providing scaffolding. Findings from another study indicate that scaffolding can lead to better learning outcomes for invention activities (Holmes et al., 2014). Although the present study only examined the analytical dimension of solution quality and more research is needed to evaluate learning effectiveness, the results demonstrate that our construct is suitable for assessing depth of comparison and thus analytical solution quality for invention activities and worked solutions. Further research is needed to examine whether the construct analytical solution quality can capture the learning processes of comparing contrasting cases in other domains where research is predominantly on problem-solving prior to instruction.

Overall, this study contributes to the research on task formats in problem solving prior to instruction. It has enabled the integration of the task formats invention activity and worked solution into teacher education, where there has so far been little research on the formats. It also showcases an innovative approach to evaluating comparative activities with contrasting cases during the problem-solving phase, the analytical solution quality.

References

Adick, C. (2018). Bereday and Hilker: origins of the 'four steps of comparison' model. *Comparative Education*, *54*(1), 35–48. https://doi.org/10.1080/03050068.2017.1396088

Alfieri, L., Nokes-Malach, T.J., & Schunn, C.D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Anderson, L.W. (2014). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Pearson.

Apthorp, H. (2010). Identifying similarities and differences. In A.D. Beesley, & H.S. Apthorp (Eds.), *Classroom instruction that works* (2nd ed., pp. 14–31).

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, *32*(1), 229–247. https://doi.org/10.1007/s10648-019-09500-5

Bereday, G.Z.F. (1964). *Comparative method in education*. Holt, Rinehart, & Winston.

Berliner, D.C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Berliner, D.C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, *24*(3), 200–212. https://doi.org/10.1177/0270467604265535

Blömeke, S., Gustafsson, J.-E., & Shavelson, R.J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift Für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally and socially determined. In D. Clandinin, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (pp. 783–802). SAGE Publications. https://doi.org/10.4135/9781526402042.n45

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D.C. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, *39*(3), 25–31. https://doi.org/10.1177/002248718803900306

Chaplain, R. (2008). Stress and psychological distress among trainee secondary teachers in England. *Educational Psychology*, *28*(2), 195–209.

Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. https://doi.org/10.1016/j.learninstruc.2016.06.007

Chi, M., Dohmen, I., Shemwell, J.T., Chin, D.B., Chase, C.C., & Schwartz, D.L. (2012). *Seeing the forest from the trees: A comparison of two instructional models using contrasting cases*. Annual meeting of the American Educational Research Association.

Chin, D.B., Chi, M., & Schwartz, D.L. (2016). A comparison of two methods of active learning in physics: inventing a general solution versus compare and contrast. *Instructional Science*, *44*(2), 177–195. https://doi.org/10.1007/s11251-016-9374-0

Commons, M.L., Trudeau, E.J., Stein, S.A., Richards, F.A., & Krause, S.R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, *18*(3), 237–278. https://doi.org/10.1006/drev.1998.0467

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction*, *51*, 26–35. https://doi.org/10.1016/j.learninstruc.2016.11.002
Glogger-Frey, I., Treier, A.-K., & Renkl, A. (2022). How preparation-for-learning with a worked versus an open inventing problem affect subsequent learning processes in pre-service teachers. *Instructional Science*, 50, 451–473. https://doi.org/10.1007/s11251-022-09577-6
Gold, B., Hellermann, C., & Holodynski, M. (2016). Professionelle Wahrnehmung von Klassenführung – Vergleich von zwei videobasierten Erfassungsmethoden [Professional vision of classroom management – comparison of two video-based methods]. In D. Prinz, & K. Schwippert (Eds.), *Der Forschung – Der Lehre – Der Bildung* (pp. 103–118). Waxmann.
Gold, B., & Holodynski, M. (2017). Using digital video to measure the professional vision of elementary classroom management: Test validation and methodological challenges. *Computers & Education*, *107*, 13–30. https://doi.org10.1016/j.compedu.2016.12.012
Gold, B., Pfirrmann, C., & Holodynski, M. (2020). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education*, *72*(4), 431–447. https://doi.org/10.1177/0022487120963681
Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. SAGE.
Heemsoth, T., & Kleickmann, T. (2018). Learning to plan self-controlled physical education: Good vs. problematic teaching examples. *Teaching and Teacher Education*, *71*, 168–178. https://doi.org/10.1016/j.tate.2017.12.021
Hilker, F. (1962). *Vergleichende Pädagogik [Comparative education]*. Max Hueber.
Holmes, N. G., Day, J., Park, A. H. K., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, *42*(4), 523–538. https://doi.org/10.1007/s11251-013-9300-7
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88. https://doi.org/10.1016/j.tate.2013.11.004
Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. Springer US. https://doi.org/10.1007/978-1-4614-5149-5
Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1
Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, *29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x
Loibl, K., & Rummel, N. (2014). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, *42*(3), 305–326. https://doi.org/10.1007/s11251-013-9282-5
Loibl, K., Tillema, M., Rummel, N., & van Gog, T. (2020). The effect of contrasting cases during problem solving prior to and after instruction. *Instructional Science*, *48*(2), 115–136. https://doi.org/10.1007/s11251-020-09504-7

Marzano, R.J., Pickering, D.J., & Pollok, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Association for Supervision and Curriculum Development.
Mayring, P. (2015). *Qualitative Inhaltsanalyse [Qualitative content analysis]* (12th ed.). Juventa.
Plöger, W., Krepf, M., Scholl, D., & Seifert, A. (2020). Analytical competence of teachers: Assessing the construct validity by means of mixed methods and drawing consequences for teacher education. *Teacher Education Quarterly*, *47*(2), 134–158.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086
Schipolowski, S., & Böhme, K. (2016). Assessment of writing ability in secondary education: comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1 Educational Studies in Language and Literature*, *16*, 1–22. https://doi.org/10.17239/L1ESLL-2016.16.01.03
Schwartz, D.L., Chase, C.C., Oppezzo, M.A., & Chin, D.B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140
Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321
Sherin, M.G. (2007). The development of teachers´ professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S.J. Denny (Eds.), *Video research in the learning sciences* (pp. 383–395). Lawrence Erlbaum Associates, Publishers.
Sherin, M.G., & van Es, E.A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, *60*(1), 20–37. https://doi.org/10.1177/0022487108328155
Stürmer, K., Könings, K.D., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *The British Journal of Educational Psychology*, *83*, 467–483. https://doi.org/10.1111/j.2044-8279.2012.02075.x
Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi:10.1023/a:1022193728205
Wedde, S., Busse, A., & Bosse, D. (2021). Zur Wirksamkeit von Invention Activities auf das Lernen von Lehramtsstudierenden [The effectiveness of invention activities on student teacher learning]. *Zeitschrift Für Hochschulentwicklung*, *16*(1), 35–54. https://doi.org/10.3217/zfhe-16-01/03
Wedde, S., Busse, A., & Bosse, D. (2022). Content-related solution quality in invention activities and worked solutions – promoting the professional vision of classroom management. *International Journal of Modern Education Studies*, *6*(2), 523–546.
Wiedmann, M., Leach, R.C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science*, *40*(4), 711–730. https://doi.org/10.1007/s11251-012-9204-y
Wilcke, H., & Budke, A. (2019). Comparison as a method for geography education. *Education Sciences*, *9*(3), 225. https://doi.org/10.3390/educsci9030225
Wolff, C.E., Jarodzka, H., van den Bogert, N., & Boshuizen, H.P.A. (2016). Teacher vision: expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, *44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z