*Further exploration of the classroom
video analysis (CVA) instrument as a
measure of usable knowledge for teaching
mathematics: taking a knowledge system
perspective*

**Nicole B. Kersting, Taliesin Sutton,
Crystal Kalinec-Craig, Kathleen Jablon
Stoehr, Saeideh Heshmati, Guadalupe
Lozano, et al.**

ZDM

Mathematics Education

ISSN 1863-9690

ZDM Mathematics Education

DOI 10.1007/s11858-015-0733-0

Your article is protected by copyright and all rights are held exclusively by FIZ Karlsruhe. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Further exploration of the classroom video analysis (CVA) instrument as a measure of usable knowledge for teaching mathematics: taking a knowledge system perspective

Nicole B. Kersting¹ · Taliesin Sutton¹ · Crystal Kalinec-Craig¹ ·
Kathleen Jablon Stoehr¹ · Saeideh Heshmati¹ · Guadalupe Lozano¹ ·
James W. Stigler²

Accepted: 4 September 2015
© FIZ Karlsruhe 2015

Abstract In this article we report further explorations of the classroom video analysis instrument (CVA), a measure of usable teacher knowledge based on scoring teachers' written analyses of classroom video clips. Like other researchers, our work thus far has attempted to identify and measure separable components of teacher knowledge. In this study we take a different approach, viewing teacher knowledge as a system in which different knowledge components are flexibly brought to bear on specific teaching situations. We explore this idea through a series of exploratory factor analyses of teachers' clip level scores across three different CVA scales (fractions, ratio and proportions, and variables, expressions, and equations), finding that a single dominant dimension explained from 55 to 63 % of variance in the scores. We interpret these results as consistent with a view that usable teacher knowledge requires both individual knowledge components, and an overarching ability to access and apply those components that are most relevant to a particular teaching episode.

1 Introduction

Understanding what mathematics teachers need to know, and what it takes to be able to apply that knowledge in the classroom, is critical for helping teachers improve their practice and their students' learning. For years, progress toward this goal was hampered by imprecise

and inconsistent use of terminology, a lack of well-developed theories, and a paucity of measures (Baumert & Kunter, 2013). Despite recent progress on all of these fronts, (Ball, Thames, & Phelps, 2008; Baumert & Kunter, 2013; Hill & Ball, 2004; König, Blömeke, & Kaiser, 2015; Seidel, Blomberg, & Stürmer, 2010; Stürmer & Seidel, 2014), we still know little about how the knowledge teachers acquire becomes usable, and how teachers apply it in the process of teaching (Ball et al., 2008). In this paper, we report on our own instrument development effort, the classroom video analysis (CVA) instrument, which is based on teachers' ability to analyze teaching events (Kersting, 2008; Kersting, Givvin, Sotelo, & Stigler, 2010; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Kersting, Sherin, & Stigler, 2014). Specifically, we explore what we can learn from teachers' scores on the CVA about usable knowledge and knowledge use.

1.1 Advancements in research on teacher knowledge in mathematics

Recent contributions to the study of teacher knowledge have been made in three important areas: (a) Identifying knowledge domains that are relevant for teaching, (b) developing instruments to measure these knowledge domains (thus making it possible to conduct systematic empirical studies), and (c) using video in item design to assess teachers' ability to apply knowledge in a context that is closer to a real teaching situation.

Over the last decades, research on teacher knowledge in the area of mathematics has increasingly focused on the classroom as the place where teachers need to apply their knowledge in order to make instructional decisions that support student learning (Ball, 2000; Ball & Bass, 2000;

✉ Nicole B. Kersting
nickik@email.arizona.edu

¹ University of Arizona, Tucson, AZ, USA

² University of California, Los Angeles, CA, USA

Baumert & Kunter, 2013; Kersting et al., 2010; 2012; Shulman, 1986; 1987). The work of Shulman was seminal in advancing our theories of teacher knowledge (Shulman, 1986; 1987). Shulman's notion of pedagogical content knowledge (PCK), which he envisioned to be at the intersection of teaching and learning, highlights the importance of domain-specific knowledge as a component of professional knowledge. Shulman conceptualized PCK as blending content knowledge with knowledge of learners, learning, and pedagogy (Ball & Bass, 2000; Shulman, 1986; 1987).

A number of researchers built on and successfully extended Shulman's ideas in the area of mathematics, among them Ball, Hill, and colleagues (Hill, Schilling, & Ball, 2004; Ball & Bass, 2000). Ball recognized that although PCK can provide a useful "anticipatory resource for teachers, it sometimes falls short in the dynamic interplay of content and pedagogy in teachers' real time problem solving" (Ball & Bass, 2000, p. 88). Interested in both what mathematics teachers know and in how they use their knowledge in the classroom (Ball & Cohen, 1999), Ball and colleagues identified and analyzed mathematics as it unfolds in the process of teaching. They developed the mathematics knowledge for teaching (MKT) construct, consisting of six subdomains, along with multiple-choice items to measure the domains (Ball, et al., 2008; Hill, & Ball, 2004; Hill et al., 2004). Factor analysis results indicated a dominant general factor, which they interpreted as common knowledge of content, and which suggested an influence of teachers' general grasp of mathematics on their responses to items. In addition to the general factor, specific factors reflecting different subdomains accounted for additional variance, although item loadings on subdomains were less consistent (Hill et al., 2004). Subsequently, Ball and colleagues were able to show, at least in some studies, that teachers' scores on the MKT were related to instructional quality and student learning (Hill et al., 2008; Hill, Rowan, & Ball, 2005). Although empirical evidence suggests that the MKT represents a relevant professional knowledge domain, Ball and colleagues (2008) note that "how such knowledge is actually used and what features of pedagogical thinking shape its use, remains tacit and unexamined" (p. 403).

Another interesting set of findings and contributions comes out of the COACTIVE project. As part of the COACTIVE project, Baumert and colleagues developed a comprehensive model of Teacher Professional Competence, which along with beliefs, motivation and self-regulation, includes professional knowledge as a

key component (Baumert & Kunter, 2013). To be able to test their model empirically, Baumert and colleagues developed paper-and-pencil measures that could assess each knowledge domain separately and examined the hypothesized relationships among the different knowledge domains. Among the five domains of professional knowledge, they identified three closely related to teaching: Content knowledge, pedagogical content knowledge, and pedagogical/psychological knowledge. Baumert and colleagues (2010) were able to show that pedagogical content knowledge could be distinguished empirically from content knowledge, and that pedagogical content knowledge had a substantial positive effect on student learning that was mediated by cognitive activation and individual learning resources. Being able to study the interplay between different knowledge domains and how they develop in connection with each other represents an important step toward understanding knowledge growth.

Finally, a number of instrument development efforts, including our own, have used classroom video in the item design to investigate the application of knowledge in a context that is closer to the classroom and actual teaching performance. These studies recognize that the process of teaching relies on teachers' ongoing interpretation of classroom events, which inform instructional decisions, and that teachers who are better in interpreting teaching situations are more likely to make more informed decisions toward a specific instructional goal than teachers who are less skillful. This premise is supported by findings from the literature on expertise. Such studies have identified systematic differences in the way expert and novice teachers perceive and interpret classroom instruction, and have concluded that some of these differences can be explained by differences in knowledge (Berliner 1989; 1994; 2001; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987). Although these studies did not aim to understand in detail how differences in expert and novice performance relate back to differences in knowledge, there seems to be some agreement that the knowledge of expert teachers is organized and structured differently from that of novices. These ideas are also reflected in the concept of teacher noticing (Sherin & Hahn, 2004; Sherin & van Es, 2005; 2009; van Es & Sherin, 2008), hypothesized to be not only an indicator of teacher expertise, but also a possible mechanism for developing expertise. Being able to analyze videos of teaching events carries over into teachers' analysis of their own practice, thus creating the conditions for reflection and learning that are not unlike what

the expertise literature describes as deliberate practice (Ericsson, Krampe & Tesch-Romer, 1993).

A related line of inquiry is the work on professional vision that has come out of the LUV project (Stürmer & Seidel, 2014). To assess pre-service teachers' professional vision, Seidel and colleagues developed the video-based Observer Research Tool (Seidel et al., 2010) and eventually the Observer (Extended) Research Tool (Stürmer & Seidel, 2014), which combines video vignettes with rating scale items. Seidel and colleagues (Stürmer & Seidel, 2014) have shown that their instrument reliably assesses pre-service teachers' ability to reason about classroom events depicted on video (i.e., describing, explaining, and predicting) as they evaluate such things as goal clarity, teacher support, and learning climate. An exploratory study of pre-service teachers learning over 2 years indicated that pre-service teachers' professional vision changed over the course of the program (Stürmer, König, & Seidel, 2013). Similarly, Yeh and Santagata reported in a study of pre-service teachers who had taken a modified methods course integrating analysis skills, that those who had taken the modified methods course improved in their ability to analyze teaching events depicted in video clips by using student evidence in their analyses, compared to pre-service teachers who had taken the regular course (Yeh and Santagata 2015).

Another video-based measure to assess teachers' professional competence was developed in the context of the TED-FU study (König et al., 2014), a follow-up study to the international TED-M study, which compared teacher preparation in 17 different countries. The measure, which consists of three video vignettes, assesses teachers' ability to perceive and interpret classroom events through a number of higher- and lower-inference rating scale items attached to each of the vignettes (König, Blömeke & Kaiser, 2015). Findings showed that after 3 years in the classroom, German middle school mathematics teachers' ability to perceive and interpret teaching situations varied systematically in relation to the proportion of teaching time to overall work time, a measure the researchers interpreted as an indicator of deliberate practice (König et al., 2015).

Together, these studies demonstrate the progress the field has made in defining the different kinds of knowledge that are relevant for teaching, and in developing measures to study how these different kinds of knowledge relate to each other, to teaching and student learning, and to other variables of interest. Still missing are studies that help us better understand how knowledge gets activated and used in classroom settings. It is within this broader context that our own work is situated.

1.2 Our prior work

In our own instrument development efforts, we explicitly aimed to design a measure to capture the knowledge that teachers are able to activate and use in a classroom situation. We reasoned that even though teachers might have a lot of different knowledge, the knowledge most likely to affect teaching and student learning will be the knowledge they can access and apply in the classroom. In our measure, which we call the classroom video analysis instrument (CVA) we present teachers with video clips taken from real mathematics classrooms in order to roughly approximate a real classroom situation (Kersting, 2008). We ask teachers to view the video clips online, and to submit written analyses of what they see. Teachers' responses are scored according to four rubrics, generating measures of skills that seem basic to the work of teaching: analyzing the mathematical content and student thinking, generating suggestions for improving the teaching episode, and interpreting the teaching episode in depth. We hypothesized that the knowledge teachers' are able to access and use in their written analyses of our video clips would also be available to them in real teaching situations.

In our work to date we have developed CVA scales for three different topic areas, each based on its own set of video clips: Fractions (F), ratio and proportions (RP), and variables, expressions and equations (VEE). For all three CVA scales we have found that teachers' scored responses to the video clips are positively and strongly correlated with their scores on the MKT (Kersting et al., 2014). Further, in work focused on just the fraction scale, we have found that teachers' total scores on the CVA, as well as subscores on each of the four rubrics separately, strongly relate to instructional quality. Perhaps surprisingly, we found that one of the rubrics, suggestions for improvement, directly predicted student learning, while we observed indirect effects, mediated by instructional quality, for the remaining rubrics (Kersting et al., 2010; 2012).

We also investigated the structure underlying the relationships among teachers' scored responses. Confirmatory factor analysis of the individual rubric scores assigned to teachers' responses indicated that relationships between CVA scores were best explained by four strongly related factors, which corresponded to the four scoring rubrics. The results suggested multidimensionality and indicated that clustering of scores within each rubric was stronger than clustering across all rubric scores. Nevertheless, the analyses also showed that for practical purposes, a solution based on a single

underlying factor was reasonable, which suggested that much of the correlation between the four rubrics represented commonly shared variance among all scores. Considering the empirical results from these initial studies, we hypothesized that teachers' scores on the CVA might reflect distinct, yet closely related dimensions of usable knowledge.

1.3 Current study

In our analyses so far we have focused on individual rubric scores to understand how our scoring rubrics function across clips and how the rubrics relate to each other and other variables of interest. What we have not done yet is to analyze teachers' clip-level scores. Although we assign each teacher four scores for each video clip, it is important to remember that the scores are all constructed based on a single open response. Because teachers will only write so much, and because they don't know how their responses are being scored, it seems reasonable that they would focus on the kind of analysis most relevant for the clip at hand. Thus, for one clip they might focus on student thinking, but for another on a suggestion for improvement. This led us to wonder if we might get a better indicator of teachers' knowledge by summing the four rubric scores for each clip. Perhaps what we want to understand is not the separate knowledge components, but the degree to which teachers' are able to flexibly and strategically access these components in real time (Alexander & Judy, 1988). In this sense, perhaps teachers' knowledge is best thought of as a system designed to produce the most useful analysis of each clip.

Thus, in this study we analyze teachers' aggregate clip level scores from three different CVA scales (F, RP, and VEE) to see what they might tell us about the functioning of teachers' knowledge as a system. If we view teachers' clip level scores as indicators of teachers' ability to flexibly and strategically access different knowledge components, we might expect a single factor solution. To understand the structure underlying teachers' clip level scores we factor analyzed these scores using exploratory procedures for each of the three CVA scales.

Interpretation of our results, however, will need to take into careful consideration what we know about the effects of analyzing aggregate scores on dimensionality. Analytically and based on literature on item parceling (Bandalos & Finney, 2009; Magnus, 2013), we expect that creating aggregate scores by summing the individual rubric scores for each clip will reduce

the number of factors obtained from the analysis of individual rubric scores, unless clip characteristics or content facets produce considerable clustering resulting in new and different factors (Bandalos & Finney, 2009; Magnus, 2013). Similarly, we expect that factor loadings estimated for the aggregate scores will be larger than those observed in the analyses of individual rubric scores (Bandalos & Finney, 2009; Magnus, 2013). Hence, being able to interpret the factor analysis results of the clip level scores in a meaningful way rests on the assumption that the ability or knowledge underlying teachers' analyses of the teaching episodes, when considered as a whole, is different than the knowledge reflected in the individual rubric scores. If this argument can be made convincingly, then clip level scores are not simply aggregates and the factor analytic results do not represent a statistical artifact, but have a meaning of their own and are valid indicators of an underlying ability.

2 Methods

2.1 The classroom video analysis (CVA) instrument

The CVA instrument, which is based on teachers' ability to analyze authentic teaching events, is designed to measure the kind of knowledge that teachers can access and apply in the classroom. The approach builds on findings from research on expertise that has shown that expert and novice teachers perceive and interpret classroom events differently, which has been linked at least in part to differences in their knowledge (Carter et al., 1988; Carter et al., 1987).

To approximate as much as possible a real teaching situation in which to elicit their knowledge, teachers view short, mathematically and pedagogically interesting video clips of authentic classroom instruction online and comment in writing on 'how the teacher and the student(s) interact around the mathematical content' (Cohen, Raudenbusch, & Ball, 2003). We intended the prompt, which is the same for all video clips, to provide some focus for teacher responses by mentioning the teacher, the student, and the content. At the same time, we purposefully kept the wording broad because we expected that teachers with different levels of knowledge would focus on different aspects of the teaching episodes in their written responses.

The video clips are each between 1 and 3 min long and feature student mistakes, teacher assistance episodes, student questions and the ensuing discussion,

or interesting teaching strategies or moves to provide a rich stimulus for teachers' analyses. In addition, we select video clips in such way that they cover as much as possible important mathematics ideas within a given content area. Even though there is no rewind button in a real classroom situation, we allow teachers to view a clip more than once if they want, compensating in part for the fact that teachers are unable to interact with and probe students' thinking in a video as they would in a real classroom.

To obtain measures of teachers' knowledge, each response is scored according to four rubrics that reflect common teaching tasks. We rate the degree to which a response analyzed the mathematics shown in the video clip (MC), and student thinking or understanding (ST), the degree to which a response included suggestions for improvement (SI) and we rated the overall interpretative depth and coherence of the response. Each of the four rubrics consists of three ordered categories (0–2). To obtain teachers' clip level scores, we summed the individual rubric scores for each clip response and hence clip level scores can range from 0 to 8.

For the mathematical content (MC) rubric we assigned a score of 0, if a response did not address the mathematics shown in the video clip, a score of 1 if the mathematics or mathematical problem in the video clip was addressed descriptively but not further analyzed, and a score of 2 if the mathematics was analyzed beyond what was observable in the video clip. A score of 0 on the student thinking rubric (ST) was assigned, if a response did not address student thinking or understanding, a score of 1 was assigned if there was some concern for student thinking or understanding without analyzing it in the context of the specific mathematics, and a response obtained a score of 2 if student thinking or understanding was analyzed in explicit connection to the mathematics shown in the clip.

For the suggestions for improvement (SI) rubric, a response received a score of 0 if it did not contain any suggestion for improvement, it received a score of 1 if it included a general pedagogical suggestion and a score of 2 if the suggestion was mathematically based or directly related to the mathematics shown in the video clip. Finally, we scored responses that contained no interpretations or substantiated judgments as 0 on the depth of interpretation (DI) rubric. Responses that contained some interpretation or substantiated judgments, but did not connect the different analytic points, were scored as 1, while responses in which different interpretative points were connected to form a coherent argument were scored as 2. It is important to note that

under the DI rubric credit can be given to general pedagogical observations that are not captured under the previous three rubrics as long as they represent interpretations or substantiated judgments. In that way the DI rubric is independent from the other rubrics (it is possible although infrequent to obtain a score of 2 on the DI rubric while obtaining scores of 0 on the remaining 3 rubrics), but parts of the response that received scores under other rubrics are considered to evaluate the overall depth and coherence of the response.

Because we score each teacher response with four rubrics it is important to consider whether the rubrics capture redundant aspects of knowledge, especially when we sum the individual rubric scores to obtain a total score for a given clip as we did in this study. To avoid redundancy, we constructed the rubrics in such way that they can be linked to a unique text portion in the response. For example, if, in a given response, student thinking was analyzed in terms of the mathematics, then a score of 2 on the ST dimension would be linked to that portion of the response. If in that same response, the analysis of student thinking led to a general pedagogical suggestion, the suggestion would receive a score of 1 on the SI rubric because the suggestion itself was not mathematical. Of course, it is also possible that a response contains a mathematical analysis of student thinking and a mathematically based suggestion for improvement, and hence such a response would receive a score of 2 for both rubrics.

Scored example responses from the RP and VEE scales are shown in Table 1. To illustrate differences between teachers' responses both with regard to the knowledge captured by the individual rubric scores and with regard to teachers' ability to access and flexibly combine different knowledge, reflected in the clip level scores, we describe the two example responses for the teaching episode about patterns (VEE) in more detail.

The first VEE response has two different and somewhat unconnected foci. The first part of the response addresses the small class size and its affordances for teaching and learning in terms of general teaching strategies and reflects aspects of pedagogical/psychological knowledge. In the second part of the response the focus is on the mathematics. Although the response does not directly state the functional rule " $P(n) = 2n$, $n = \text{row number}$ ") to describe the pattern, the teacher appears to be aware of it by drawing connections to predicting values, linear equations, and multiplicative relationships, for which the response receives a score of 2 on the MC rubric. The response, however, reveals little about how these ideas are connected, or why

Table 1 Scored example responses for the RP and VEE CVA scales

RP	VEE
<p><i>Clip description:</i> students are learning about the meaning of part to part and part to whole ratios. They are working with red and yellow chips. One student ends up with 8 red and 0 yellow chips after 8 draws and concludes that his ratio represents part to whole</p> <p><i>Example response:</i> the student had a somewhat confusing example, because of the 0 yellow tosses. The teacher got the student to say that it was a part to part relationship, but he was mostly listening and guessing what she wanted to hear, as opposed to further, deeper mathematical thinking</p>	<p><i>Clip description:</i> in this clip students are working on patterns. The teacher is going over a problem which students had worked in pairs. The problem stated that a garden had 2 plants in the 1st row, 4 plants in the 2nd row, 6 plants in the 3rd row, and so on. The task was to determine how many plants were in the 10th row. To help students figure this out they had been given chips to represent the plants</p> <p><i>Example response:</i> one of the things that gets my attention in this video is the size of the class. It doesn't seem like there are many students in this class, which can make a significant difference in the lesson. If a teacher only has 8 students or so, much can be done and there can be more individualized attention</p> <p>The students are making a connection to linear equations by using objects. It seems like they understand what's going on, and the teacher reinforces what they already know in terms of multiplication and predicting future values</p> <p><i>Example response:</i> the teacher goes too quickly from "add two" to "multiply by two". It seems as if she wants students to understand that they can figure out a rule without knowing every row before that row, but she doesn't make the distinction between these two. There is a difference between the recursive rule (add to the previous term) and the functional rule (use the term number). The teacher goes on to say that this works because multiplying is the same as repeated addition. This works in this problem because row 1 had 2 plants. I hope the next problem the students do works like this so the teacher can distinguish between the two types of rules</p>
<p><i>Example response:</i> the student was not asked why he thought the ratio was part-to-whole. It was good that the teacher had him make the 8 red to 0 yellow with his coins so he could actually see what they were discussing. When the teacher asked the student if they looked at the whole group yet, there was no response from the student. She just said, "so that wouldn't make it a part to whole, it would make it a..." Essentially, she told the student the correct answer</p> <p>I'm sure what was confusing the student on this problem was the fact that the 8 red coins WERE the entire set of red coins, which the student saw as the whole. It would have been good for the teacher to point this out, but note that in the ratios that he had written (correctly!), he was still comparing two parts of the same data set</p>	<p>Individual rubric scores and clip total score</p> <p>MC: 2 ST: 1 SI: 0 DI: 1 Clip total: 4</p> <p>Individual rubric scores and clip total score</p> <p>MC: 1 ST: 1 SI: 0 DI: 1 Clip total: 3</p> <p>Individual rubric scores and clip total score</p> <p>MC: 1 ST: 2 SI: 2 DI: 2 Clip total: 7</p>

understanding those connections might be important mathematically and for student learning. We might say that this portion of the response reflects content knowledge. Even though the response does not analyze students' mathematical thinking or understanding directly, the response makes some broad claims about student understanding of patterns without specific evidence, which is reflected in a score of 1 on the ST rubric. The response does not include a suggestion for improvement and hence receives a score of 0 on the SI rubric. Finally, the response receives a score of 1 on the DI rubric because it does offer interpretations, but the two main points appear unconnected. The individual rubric scores reflect the relative strength of this teacher's content knowledge. If we assume that the entire response reflects this teacher's most meaningful interpretation of the observed teaching episode, then the clip total score of "4" might indicate an average ability to access and strategically combine different knowledge for analyzing this teaching episode.

The second VEE example response presents a more coherent analysis of the teaching episode, which earns it a score of 2 on the DI rubric. There is an immediate focus on the key mathematical ideas, the distinction between the recursive description ("It's +2") and the functional rule (" $P(n) = 2n$ ", n = row number) for describing patterns from a student learning perspective. It is clear from the response that the teacher has a solid understanding of both approaches, reflecting content and pedagogical content knowledge, which leads to a score of 2 on the MC rubric. Although he seems to wonder whether the students understood that the row number can be used to determine the number of plants, the response does not explicitly analyze students' mathematical thinking, which is reflected in a score of 1 on the ST rubric. The teacher who produced this response is able to use his understanding of repeated addition as multiplication to identify that the connection the teacher in the video draws between these two operations only works for describing some patterns, as is the case for the pattern presented in this mathematical problem, but not others. This concern leads him to conclude that it will be important in future lessons to present different kinds of pattern problems so that students are able to understand this important distinction ($SI = 2$). We may call much of this teacher's demonstrated knowledge part of content knowledge and mathematics knowledge for teaching or in Shulman's terms, pedagogical content knowledge. The high individual rubric scores reflect that this teacher was able to use his content and pedagogical content knowledge for specific skills, such as

analyzing the mathematics, while the high clip total score of "7" reflects this teacher's ability to activate and flexibly combine those knowledge components that are most relevant to produce a highly useful analysis of the teaching episode.

We cannot know from these VEE responses alone whether both teachers have the same knowledge base or not. Their responses, however, illustrate that each teacher drew on different knowledge to produce their analyses and that both teachers paid attention to somewhat different aspects of the teaching episode. We suggest that differences in the responses represent differences in teachers' knowledge systems, and that those differences are not captured by the individual rubric scores. The individual rubric scores are helpful for understanding whether teachers are able to use their knowledge for specific skills (e.g., analyzing student thinking, the mathematical content, or for generating suggestions for improvement) and whether they can do so across different teaching situations. The individual rubric scores cannot reflect teachers' overarching ability to activate and flexibly combine those knowledge components that produce the most useful analysis of a given teaching episode. This ability is better captured by clip total scores. From this perspective, differences in clip total scores can be interpreted as differences in teachers' knowledge systems, which might differ in terms of both the individual knowledge components and how they are organized into a domain relevant structure and in teachers' ability to effectively access and combine them. For the two example responses we might say that the second response is overall more useful and on target and reflects a more advanced knowledge system than the first response, especially, when considered for instructional decision making. In fact, if the second teacher's analysis is any indication, in a comparable classroom situation this teacher might give students a pattern problem next that would allow them to better understand the distinction between the recursive description and the functional rule.

We do not suggest that analyzing video clips of authentic mathematics instruction is the same as making sense of teaching situations in the classroom in real time. We do, however, suggest that teachers who are less skillful in analyzing the teaching situations depicted in the video clips are not likely to be more skillful analyzing teaching situations when faced with the complexity of a real classroom. In this way, the CVA might serve as a good upper bound proxy measure of the knowledge teachers can apply in a real teaching situation.

At the same time, it is important to recognize that teachers' responses to the video clips depend on their interpretation of the analysis prompt. Differences in the understanding of the prompt will affect which knowledge teachers activate and how they use that knowledge in their analyses. Finally, a lack of motivation or concentration, much like in actual teaching performance, might result in teacher responses that are poor indicators of their actual knowledge and their ability to apply it. Thus, scores on the CVA can only provide information on teachers' knowledge for teaching mathematics as demonstrated in their responses to the CVA video clips.

2.2 Analytical approach and statistical models

In previous studies we have factor analyzed teachers' individual rubric scores to understand how our scoring rubrics function and how they relate to each other. Analyzing the individual rubric scores is comparable to an item-level analysis of the different knowledge components, which is recommended for instrument development efforts, especially if the dimensionality of the instrument is not known, as was the case for the CVA (Bandalos & Finney, 2009; Magnus, 2013).

In the current study, we factor analyzed teachers' clip level scores by summing individual rubric scores for each clip. Creating aggregate scores from sets of items, also referred to as item parceling, and analyzing those aggregate scores has become an analytic strategy within structural equation modeling for a number of technical reasons, such as to increase score reliability, to reduce the number of model parameters to be estimated, and to improve parameter estimates. There exists an extensive literature around the effects of analyzing aggregate scores.

Relevant for our study is the fact that analytical proof as well as simulation and empirical studies suggest that aggregating across items or, as in our case, individual rubric scores, will affect the dimensionality of the data (i.e., change the factor structure) if the original items indicated a multidimensional structure. The extent of the effect item parceling has on dimensionality depends on the amount of multidimensionality in the original data and whether items that are more similar or less alike are combined. The main concern of those who advise against this practice is that the factor structure based on aggregate scores is difficult to interpret because it represents the ability or knowledge underlying the aggregate scores, which may not be the same as the original factors. This concern has merit especially when the results from the aggregate scores are used to interpret and label the underlying trait, the

factor structure of the individual items is not known, and when items are combined that measure quite different domains.

In the case of the CVA, however, we propose that a teacher's total clip score represents a teacher's ability to activate and strategically combine different knowledge components to produce the most useful analysis of a given teaching episode, while the individual rubric scores, which reflect knowledge components, contribute to the overall analysis but separately cannot create the same meaning as created by the analysis as a whole. In this paper we make the argument that teachers' clip level scores reflect a different ability, one that is not captured by the individual rubrics scores. Factor analyzing the clip level scores will reveal the structure of this underlying ability.

To investigate the structure of the clip level scores, we fit simple exploratory factor models to the CVA assessment data, using maximum likelihood estimation.

2.3 Data sources and description

We analyzed responses from elementary and middle school mathematics teachers to three different CVA scales (on fractions, on ratios and proportions, and on variables, expressions, and equations). All three samples were convenience samples, but were obtained based on recruiting efforts at the national level and hence should represent a considerable range of teacher backgrounds, experiences, and teaching contexts.

We analyzed scored responses from 256 teachers for the topic of fractions, 212 responses from the VEE scale and 208 responses to the CVA ratio and proportions assessment. The CVA fraction and RP scales consisted of 13 video clips each, the VEE scale comprised 14 clips in total. The CVA fraction scale contained clips that addressed the meaning of fractions, the idea of equivalence, comparing fractions, and all four fraction operations. The ratio and proportion scale consisted of video clips that addressed the meaning of ratios, interpreting ratios, multiplicative reasoning, solving proportions, relationships between solving proportions and algebraic reasoning, ratios as fractions and combining ratios. Video clips that form the variables, expressions, and equation scale address the meaning of variables, solving equations, modeling with variables, understanding patterns, meaning and writing of expressions. Within each CVA scale, there was a fair amount of variation with respect to the teaching situations and mathematical ideas.

Overall, clip mean scores showed some variation within scales as shown in Table 2. The largest differences were observed for the fraction scale; average

Table 2 Means and standard deviations of total scores by clip and by topic area

# Video clip	Variables expressions and equations (VEE)		Fractions (F)		Ratios and proportions (RP)	
	M	SD	M	SD	M	SD
Clip 1 total	2.08	1.99	2.41	2.10	2.41	2.12
Clip 2 total	1.94	1.86	3.09	2.51	1.78	1.83
Clip 3 total	1.86	2.01	3.16	2.73	1.91	1.75
Clip 4 total	2.36	2.15	3.04	2.17	2.34	2.19
Clip 5 total	1.58	1.73	3.61	2.51	2.10	2.01
Clip 6 total	2.18	2.01	3.63	3.36	2.25	1.85
Clip 7 total	1.90	1.91	2.43	1.95	2.27	2.10
Clip 8 total	1.97	2.11	1.77	2.30	2.94	2.38
Clip 9 total	2.45	2.06	2.50	2.79	2.05	2.20
Clip 10 total	2.25	2.11	2.59	2.68	2.16	2.06
Clip 11 total	1.96	1.73	3.18	2.42	2.49	2.05
Clip 12 total	2.39	2.12	3.40	2.68	2.00	1.72
Clip 13 total	2.25	2.10	2.85	2.88	2.27	2.13
Clip14 total	2.23	2.22				
Mean clip score	2.1		2.90		2.22	

$$N_{VEE} = 201; N_{RP} = 212; N_F = 256$$

Table 3 Model summary table of EFA by topic

	Eigenvalue >1	% Variance	Cumulative % variance
Variables, expressions, and equations	8.820	63.00	63.00
Fractions	7.193	55.33	55.33
Ratio and proportions	7.131	54.86	54.86

Extraction method = maximum Likelihood

Table 4 Standardized factor loadings by CVA scale

# Video clips	Variables expressions and equations (VEE)	Fractions (F)	Ratios and proportions (RP)
Clip 1 total	0.789	0.698	0.719
Clip 2 total	0.772	0.734	0.698
Clip 4 total	0.815	0.680	0.697
Clip 5 total	0.841	0.840	0.785
Clip 6 total	0.750	0.854	0.744
Clip 7 total	0.845	0.840	0.688
Clip 8 total	0.824	0.801	0.737
Clip 9 total	0.816	0.627	0.778
Clip 10 total	0.762	0.713	0.788
Clip 11 total	0.812	0.719	0.767
Clip 12 total	0.757	0.726	0.766
Clip 13 total	0.759	0.750	0.716
Clip 14 total	0.821	–	–

clip total scores ranged from 1.77 to 3.63. Variation in average clip scores for VEE and RP were smaller, ranging from 1.58 to 2.45, and 1.78–2.94, respectively. The observed variation in average clip scores might indicate that some clips were easier to analyze than others, or that some clips offered more opportunities for analysis than others. Standard deviations also varied within scales, indicating that some clips produced greater variation in responses than others. Finally, the distribution of clip level total scores was somewhat skewed with more responses receiving a total clip scores of 0, 1, or 2 than receiving clips scores of 3 and higher. The mean clip total scores, averaged across all clips for each scale, were 2.1 and 2.2 for VEE and RP respectively, and 2.9 for fractions. The comparatively low values could reflect characteristics of our samples or overall difficulty of the scale, or both. The variation in average scores across scales is equally difficult to interpret because it might reflect differences in teacher ability across samples or that the teaching episodes shown in the video clips were relatively more difficult for one topic area than another, or both.

3 Results

Across all three CVA scales, exploratory factor analysis results indicate a strong, single dimension that explains a considerable proportion of variance in teachers' total clip scores. For the variables, expressions, and

equations CVA scale, a single factor explained 63 % of the variance, and for the fraction and ratio and proportion scales, a single factor explained 55 % of variance, respectively (Table 3). Additional eigenvalues reflecting additional shared variance among total clips scores (not already explained by the first factor), were below the commonly used Kaiser cutoff values of 1.0 (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kaiser, 1960), and hence negligible. Scree plots of the Eigenvalues are presented in Fig. 1a–c in the “Appendix”.

Factor loadings based on total clip scores were large and fairly consistent across clips and scales as shown in Table 4. The standardized loadings range from .63 to .85, which can be interpreted as representing the correlation between the item and the underlying factor. The largest range of factor loadings is observed for the fraction scale, the least variation for the ratio and proportion scale. The results suggest that teachers’ clip level scores are good indicators of the ability to access and strategically combine knowledge to produce the most useful analysis of a given teaching episode.

4 Discussion

In this study we further explored the classroom video analysis instrument, which is based on teachers’ ability to analyze teaching events shown in short video clips of authentic mathematics instruction, as a measure of teachers’ usable knowledge. In our prior work, we, like other researchers, have attempted to identify and measure separable components of teacher knowledge. In the current study we took a different approach, viewing teacher knowledge as a system in which different knowledge components are flexibly brought to bear on specific teaching situations. To explore this idea we carried out a series of exploratory factor analyses using clip total scores, which we computed by summing the individual rubric scores for each clip, from three different CVA scales (fractions, ratio and proportions, and variables, expressions, and equations).

Results from our exploratory factor analyses indicated a single, strong dimension underlying teachers’ clip level scores for all three CVA scales, which explained between 55 and 63 % of the variance. Factor loadings were large across the board (ranging from .62 and .85) and suggest that clip level scores of teachers’ analyses of each teaching episode are good indicators of the single underlying dimension.

We propose that this single dimension represents usable knowledge, which we conceptualize here as a knowledge system, and which expands the notion of usable knowledge we suggested in our earlier work. In

previous studies we hypothesized that the individual rubric scores captured different dimensions of teachers’ usable knowledge because teachers were able to access and draw on different kinds of knowledge when they analyzed the observed teaching episodes. This interpretation was supported by the multidimensionality we found when we factor analyzed the individual rubric scores. For the current study, in which we analyzed teachers’ entire responses to each teaching episode by using clip total scores, we propose that usable knowledge requires both individual knowledge components, and an overarching ability to access and apply those components that are most relevant to a particular teaching episode. We hypothesize that teachers’ responses do not represent all the knowledge teachers have and could bring to bear on a particular teaching situation, but rather the subset of knowledge that teachers deem most relevant and essential. In that way, teachers’ clip level scores represent a functional aspect of knowledge and its application that is not captured by the individual rubric scores. If a teacher’s analysis of a teaching episode as a whole can be considered an output of his knowledge system, then qualitative differences in teachers’ analyses can be interpreted as differences in teachers’ knowledge systems and can be quantified to reflect differences in their usable knowledge.

One advantage of taking a knowledge system perspective is that it captures well the dynamic and fluid nature of knowledge, knowledge use, and knowledge growth. Being able to draw on the most relevant knowledge for a particular situation is not independent from the individual knowledge components a teacher has and how these components are organized, but rather depends on and interacts with them. Under a system view, greater usable knowledge can mean more individual domain-relevant knowledge components, but it also means having knowledge organized in a structure that reflects meaningful relationships between the components according to domain-specific content, tasks, and uses. It also means being better at accessing the most relevant or useful knowledge for a given situation, or for a particular purpose. An added benefit of taking a system perspective is that knowledge growth can be imagined as a non-linear process of adaption, revision, and evolution that can occur in different areas and at different levels of the system or combination of levels, by acquiring new individual knowledge components, by improving the knowledge structure, and by improving the ability to access and flexibly use the knowledge that is most relevant and useful for a specific context.

Assessments of teacher knowledge in mathematics have primarily focused on measuring different components of relevant knowledge. So far less attention has

been paid to measuring functional aspects of knowledge and knowledge use. To better understand how teachers develop knowledge over time and how teachers use their knowledge in the process of teaching, we might need instruments that can measure both. Without further study it is impossible to say whether teachers' clip total scores can provide measures of functional aspects of teacher knowledge, but at the very least, the CVA might help us think about how to design such instruments.

The ability to activate and use the most relevant knowledge for interpreting each observed teaching episode, is related to what studies on expertise have found for expert chess players. It is not that expert chess players are superior to novice or intermediate chess players because they are able to predict more turns ahead in the game but because they are able to identify for any turn a highly effective set of possible moves from among which they select the best possible move for a specific situation (National Research Council, 2000). Functional aspects of knowledge have also been recognized in some studies on teacher cognition by distinguishing functional or procedural knowledge, which is knowledge about how to carry out domain-specific tasks as, for example, knowledge of how to assess student thinking or understanding, from factual or declarative domain-relevant knowledge (Alexander & Judy, 1988; Reynolds, Sinatra, & Jetton, 1996). To be sure, teachers' written responses on the CVA cannot help us understand the cognitive processes that led to their creation, but if taken as outputs of teachers' knowledge systems, they do provide preliminary evidence of their existence.

The two example responses we considered earlier provide some indication of differences in both teachers' knowledge systems. Although responses from both teachers did address the mathematics shown in the video clip, rules for expressing patterns, only the second response did so in a highly connected way. By carefully comparing the affordances of the recursive and functional rules for describing patterns and by contemplating the importance of this difference for student understanding, the response reveals both a highly connected domain-relevant knowledge structure and the ability to draw on the most relevant knowledge for interpreting the teaching episode. In comparison, the first response seems to reflect a less advanced knowledge system on all levels. Although this response broadly links the pattern problem to linear equations, and multiplication, it does so without detail or explanation on how and why these mathematical ideas connect and why this might be important for

student understanding. The response does not distinguish between the recursive and the functional rule and does not draw attention to this important difference. Instead the primary focus of the response is on the perceived small class size in the video clip and its affordances for teaching and learning.

At this point, we cannot say with certainty, whether teachers' clip level scores on the CVA measure usable knowledge as a functional system. The factor analyses results by themselves cannot provide sufficient evidence for this interpretation, which is one limitation of our study. Future studies of the CVA need to explicitly address alternative interpretations of our current findings. What is, however, exciting about the idea of a functional knowledge system is that it can be studied and tested in systematic ways. We can imagine studies, in which we ask teachers with different knowledge levels and expertise to interpret the teaching situations shown in the video clips and to share through speak-alouds which aspects or events in the episodes attract their attention and why. We might learn that teachers are largely unaware of the exact processes that lead to their analyses or we might discover some of the rules or thinking that governs how teachers use their knowledge in the process of teaching. We can also devise studies to test how knowledge becomes usable. We could test under which experimental learning conditions teachers become able to recognize specific instructional strategies or approaches, for example, supporting and furthering student thinking, and strategically use this knowledge in their analyses of the teaching episodes. Finally, we might be able to study knowledge growth in teachers as a function of interventions that target different levels of the knowledge system.

Despite notable progress over the past two decades, much work remains to be done. We are only at the beginning of uncovering what teachers' know and how they know so that they can use their knowledge in the process of teaching. A fundamental challenge to this work is that our theoretical advances are limited by our measures and our measures are limited by our theoretical understanding. What we can learn from the CVA about usable knowledge from a system perspective depends on whether we can gather compelling evidence beyond our factor analysis results to build a convincing argument that indeed it does.

Acknowledgments This study was supported by the Department of Education, Institute of Educational Sciences, Grant No. R305M060007. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organization or agency that provided support for this project.

Appendix

See Fig. 1.

Fig. 1 Scree Plots of the Extracted Factor Structure by Topic

References

- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58(4), 375–404. doi:[10.3102/00346543058004375](https://doi.org/10.3102/00346543058004375).
- Ball, D. L. (2000). Bridging practices: intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, 51(3), 241–247. doi:[10.1177/0022487100051003013](https://doi.org/10.1177/0022487100051003013).
- Ball, D., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport: Ablex.
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey Bass.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, 59(5), 398–407. doi:[10.1177/0022487108324554](https://doi.org/10.1177/0022487108324554).
- Bandalos, D. L., & Finney, S. J. (2009). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–297). Mahwah, NJ: Erlbaum, 269–297.

- Baumert, J. & Kunter, M. (2013). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 49–63). New York: Springer. doi:10.1007/978-1-4614-5149-5.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. doi:10.3102/0002831209345157.
- Berliner, D. C. (1989). Implications of studies of expertise in pedagogy for teacher education and evaluation. In J. Pfliderer (Ed.), *New directions for teacher assessment (Proceedings of the 1988 ETS Invitational Conference)* (pp. 39–68). Princeton: Educational Testing Service.
- Berliner, D. C. (1994). The wonder of exemplary performances. In J. N. Mangieri & C. Collins Block (Eds.), *Creating powerful thinking in teachers and students* (pp. 161–186). Forth Worth, TX: Holt, Rinehard, & Winston.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482. doi:10.1016/S0883-0355(02)00004-6.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom stimuli. *Journal of Teacher Education*, 39(3), 25–31. doi:10.1177/002248718803900306.
- Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. (1987). Processing and using information about students: a study of expert, novice, and postulant teachers. *Teaching and Teacher Education*, 3(2), 147–157. doi:10.1016/0742-051X(87)90015-1.
- Cohen, D. K., Raudenbusch, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142. doi:10.3102/01623737025002119.
- Council, National Research (Ed.). (2000). *How people learn: brain, mind, experience, and school*. Washington: National Academy Press.
- Ericsson, K. A., Krampe, R. Th, & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406.
- Es, E. A., & Sherin, M. G. (2008). Mathematics teachers "learning to notice" in the context of a video club. *Teaching and Teacher Education*, 24(2), 244–276. doi:10.1016/j.tate.2006.11.005.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <http://dx.doi.org/10.1037/1082-989X.4.3.272>
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: results from California's Mathematics Professional Development Institutes. *Journal for Research in Mathematics Education*, 35(5), 330–351.
- Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cognition and Instruction*, 26(4), 430–511. doi:10.1080/07370000802177235.
- Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406. doi:10.3102/0002831204200237.
- Hill, H., Schilling, S., & Ball, D. L. (2004). Developing measures of teacher's mathematics knowledge for teaching. *The Elementary School Journal*, 105(1), 11–30. doi:10.1086/428763.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. doi:10.1177/001316446002000116.
- Kersting, N. (2008). Using video clips as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, 68(5), 845–861. doi:10.1177/0013164407313369.
- Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61(1–2), 172–181. doi:10.1177/0022487109347875.
- Kersting, N. B., Givvin, K. B., Thompson, B., Santagata, R., & Stigler, J. (2012). Developing measures of usable knowledge: teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–590. doi:10.3102/0002831212437853.
- Kersting, N. B., Sherin, B., & Stigler, J. W. (2014). Automated scoring of teachers' open-ended responses to video prompts: bringing the classroom video analysis (CVA) assessment to scale. *Educational and Psychological Measurement*, 74(6), 950–974. doi:10.1177/0013164407313369.
- König, J., Blömeke, S., & Kaiser, G. (2015). Early career mathematics teachers general pedagogical knowledge and skills: do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education*, 13(2), 331–350. doi:10.1007/s10763-015-9618-5.
- 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. doi:10.1016/j.tate.2013.11.004.
- Magnus, B. E. (2013). the influence of parceling on the implied factor structure of multidimensional item response data. (Master's Thesis). Available from the ProQuest Dissertation and Theses database. (UMI No: 1545097).
- Reynolds, R. E., Sinatra, G. M., & Jetton, T. L. (1996). Views of Knowledge acquisition and representation: a continuum from experience centered to mind centered. *Educational Psychologist*, 31(2), 93–104.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). OBSERVE -Validation of a video-based instrument assessing professional vision. *Zeitschrift für Pädagogik, (Beiheft)*, 56, 296–306.
- Sherin, M. G., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, 20(2), 163–183. doi:10.1016/j.tate.2003.08.001.
- Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education* 13(3), 475–491. ISSN: 1059-7069.
- 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. doi:10.1177/0022487108328155.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <http://www.jstor.org/stable/1175860>.
- Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Stürmer, K., König, K., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *British Journal of Educational Psychology*, 83(3), 467–483. doi:10.1111/j.2044-8279.2012.02075.xvan.
- Stürmer, K., & Seidel, T. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771. doi:10.3102/0002831214531321.
- Yeh, C., & Santagata, R. (2015). Pre-service teachers' learning to generate evidence-based hypotheses about the impact of mathematics teaching on learning. *Journal of Teacher Education*, 66(1), 21–34. doi:10.1177/0022487114549470.