10-19-2015

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

Kathleen Jablon Stoehr
Santa Clara University, kstoehr@scu.edu

Recommended Citation

Kersting, N, .Sutton, T., Kalinec-Craig, C., Stoehr, K., Heshmati,S., Lozano, G., & Stigler, J. (2015). Further exploration of the classroom video analysis (CVA) instrument as a measure of usable knowledge for teaching mathematics: taking a knowledge system perspective. ZDM Mathematics Education.

The final publication is available at Springer via http://doi.org/10.1007/s11858-015-0733-0

Further Exploration of the Classroom Video Analysis (CVA) Instrument as a Measure of Usable Knowledge for Teaching Mathematics: Taking a Knowledge System Perspective

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.

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 teacher's 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 percent 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; Koenig, Bloemke, & Kaiser, 2014; Seidel, Blomberg, & Stürmer, 2010; Stürmer & Seidel, 2014), however, 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; 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, 2002). 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 process of teaching (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 some domains (Ball, et al., 2008; Hill, H. & Ball, 2004; Hill, Schilling, & Ball, 2004). Factor analysis results indicated a dominant general factor, which they interpreted as common knowledge of content and which suggested an influence of general grasp of mathematics on teachers' responses to items. In addition, some of the specific factors accounted for substantial proportion of the 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, they 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 Baumert and colleagues identified, three closely related to teaching: Content knowledge, pedagogical content knowledge, and pedagogical/psychological knowledge. Baumert and colleagues 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 (Baumert et al., 2010). 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. They 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 is 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 two years indicated that pre-service teachers' professional vision changed over the course of the program (Stürmer, König, & Seidel, 2013). Similarly, Yeh and Santagata, in a study of pre-service teachers who had taken a modified methods course build around video cases, reported increases their ability to analyze teaching events as compared to pre-service teachers who had taken the regular course (Yeh & Santagata, 2013).

Another video-based measure to assess teachers' professional competence was developed in the context of the TED-FU study, a follow-up study to the international TED-M study, which compared teacher preparation in 17 different countries (reference). 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 (Koenig, et al., 2014). Findings showed that after three years in the classroom, German middle school mathematics teachers' ability to perceive and interpret teaching situations varied in relation the proportion of teaching time to overall work time, a measure the researchers interpreted as an indicator of deliberate practice (Koenig et al., 2014).

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 fractions 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 very 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 (master thesis and review chapter). 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 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 at 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 one and three minutes 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 the responses are 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 and 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).

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) rubric 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 individual knowledge components measured by the rubric scores and with regard to teachers' ability to access and strategically 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.

Table 1.

Scored Example Responses for the RP and VEE CVA Scales

RP		*VEE*	
Clip Description: 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.	Individual Rubric Scores and Clip Total Score	*Clip Description*: 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.	Individual Rubric Scores and Clip Total Score
Example Response: 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.	MC: 1 ST: 1 SI: 0 DI: 1 Clip Total: 3	*Example Response*: 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. 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.	MC: 2 ST: 1 SI: 0 DI: 1 Clip Total: 4
Example Response: 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	MC:1 ST:2 SI: 2 DI:2 Clip Total: 7	*Example Response*: 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	MC:2 ST:1 SII: 2 DI: 2 Clip Total: 7

wouldn't make it a part to whole, it would make it a..." Essentially, she told the student the correct answer. I'm sure what was confusing the student on this problem was the fact that the 8 red coins WAS 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.	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.

The first 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 term of general teaching strategies, and reflects dimensions of pedagogical/psychological knowledge. The response offers little in terms of student assessment and does not discuss assessment strategies the teacher in the video clip uses or could have used. In the second part of the response the focus is on the mathematics. The response reveals that the teacher knew that the functional rule (“P(n) = 2n, n = row number) to describe patterns represents a linear equations and multiplicative relationships, for which he receives as score of 2 on the MC rubric, even though his response reveals little about how these ideas are connected, and why understanding those connections might be important for student learning. Nevertheless, we might say that it reflects content knowledge and specialized content knowledge. Because the response does not analyze students’ mathematical thinking or understanding, yet suggests some assessment of student understanding the response receives a 1 on the ST rubric. The response does not include a suggestion for improvement and hence receives a score of 0. Finally, the response receives as 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 level score represents the teacher's ability to access and strategically combine different knowledge.

The second 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 idea, 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 very clear from the response that the teacher has a solid understanding of both approaches, reflecting content and pedagogical content knowledge, and 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 the mathematical thinking, reflected in a score of 1 on the ST rubric. He is able to use his understanding of repeated addition as multiplication to identify that the connection the teacher in the video draws between repeated addition and multiplication 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. 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. Again, the rubric scores provide information about the individual knowledge components the teacher used for different aspects of his analysis, while the clip level score represents teachers' ability to access and strategically use the different components. Both responses demonstrate that teachers applied knowledge in their analyses of the teaching episodes, which by definition makes it usable.

Based on our brief analysis of the example responses, three observations may be helpful when thinking about knowledge as a functional system. First, different kinds of knowledge are required to interpret the teaching episode, as demonstrated by both responses. Second, despite our attempts to define and discriminate among different kinds of knowledge, teachers' application of knowledge "in the wild" seems less like a careful analysis based on all components, and more like a flexible zeroing in on the kind of analysis most germane to the specific situation. Third, when comparing both responses it becomes clear that the second response provides a better basis for instructional decision making than the first response. In fact, if the second teacher's concern is any indication, in a comparable classroom situation this teacher might give students a pattern problem next that would allow them to 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 able to analyze teaching events in a more meaningful way 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.

Nevertheless, it is important to recognize that teachers' responses to the video clips depend on their understanding of the analysis prompt and might not reflect all the knowledge they have. 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 models 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, 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 these 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 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 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 to 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 variation in average scores across scales is 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.

Table 2.

Means and Standard Deviations of Total Scores by Clip and by Topic Area

	Variables Expressions & Equations (VEE)		Fractions (F)		Ratios & Proportions (RP)	
# Video Clip	*M*	*SD*	*M*	*SD*	*M*	*SD*
Clip 1Total	2.08	1.99	2.41	2.10	2.41	2.12
Clip 2Total	1.94	1.86	3.09	2.51	1.78	1.83
Clip 3Total	1.86	2.01	3.16	2.73	1.91	1.75
Clip 4Total	2.36	2.15	3.04	2.17	2.34	2.19
Clip 5Total	1.58	1.73	3.61	2.51	2.10	2.01
Clip 6Total	2.18	2.01	3.63	3.36	2.25	1.85
Clip 7Total	1.90	1.91	2.43	1.95	2.27	2.10
Clip 8Total	1.97	2.11	1.77	2.30	2.94	2.38
Clip 9Total	2.45	2.06	2.50	2.79	2.05	2.20
Clip 10Total	2.25	2.11	2.59	2.68	2.16	2.06
Clip 11Total	1.96	1.73	3.18	2.42	2.49	2.05
Clip 12Total	2.39	2.12	3.40	2.68	2.00	1.72
Clip 13Total	2.25	2.10	2.85	2.88	2.27	2.13
Clip14Total	2.23	2.22				
Mean Clip Score	*2.1*		*2.90*		*2.22*	

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

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 percent of the variance, and for the fraction and ratio and proportion scales, a single factor explained 55 percent 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 value of 1.0 (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kaiser, 1960), and hence negligible. Scree plots of the Eigenvalues are presented in Figure A1a through c in the Appendix.

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

Note. Extraction method = Maximum Likelihood.

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 underlying ability to access and strategically combine knowledge to produce the most useful analysis of a given teaching episode.

Table 4.

Standardized Factor Loadings by CVA Scale

# Video Clips	Variables Expressions & Equations (VEE)	Fractions (F)	Ratios & Proportions (RP)
Clip 1 Total	.789	.698	.719
Clip 2 Total	.772	.734	.698
Clip 4 Total	.815	.680	.697
Clip 5 Total	.841	.840	.785
Clip 6 Total	.750	.854	.744
Clip 7 Total	.845	.840	.688
Clip 8 Total	.824	.801	.737
Clip 9 Total	.816	.627	.778
Clip 10 Total	.762	.713	.788
Clip 11 Total	.812	.719	.767
Clip 12 Total	.757	.726	.766
Clip 13 Total	.759	.750	.716

Clip 14 Total	.821	-	-

4 Discussion

In this article 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 classroom instruction. In our prior work, we, like other researchers, have attempted to identify and measure separable components of teacher knowledge. In this 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-level scores from three different CVA scales (fractions, ratio and proportions, and variables, expressions, and equations) to see what we can learn about the knowledge the CVA measures and about how teachers activate and use their knowledge.

Results from our exploratory factor analysis of teachers' clip level scores indicated a single, strong factor underlying teachers' scores with large factor loadings for each of the three CVA scales. The total amount of variance in teachers' clip level scores explained by the single underlying dimension ranged from 55 and 63 percent and factor loadings were large across the board (ranging between .62 and .85). The results suggest that clip level scores are good indicators of a single dimension underlying teachers' analyses of the teaching episodes. The interpretation of this dimension, however, and what we can learn about the knowledge measured by the CVA depends on the meaning of teachers' clip level scores. Several different interpretations are possible.

One possible interpretation is that teachers' clip level scores simply reflect the sum of teachers' usable knowledge, that is, the same knowledge that is captured by the individual rubric scores and that is usable because it is applied to a teaching situation. In this case, we would interpret the single factor as reflecting usable knowledge, and we would have to reconcile that the same construct has different structures depending on the level at which it is measured, which raises concern. Earlier analyses of the individual rubric scores revealed a multidimensional structure. Used in this way, the term usable knowledge, means knowledge that can be applied. Tt implies that the underlying cognitive processes for applying individual knowledge components and for strategically accessing and combining those components that are most relevant for interpreting a specific teaching situation are the same.

Another interpretation of our factor analysis results might be that either the multidimensional or the unidimensional construct structure represents the "correct" construct structure of usable knowledge, while the other solution should be considered a statistical artifact. The issue with this interpretation is that it might be difficult to decide which is more appropriate because this decision would need to be based on some meaningful rationale.

There is, however, a third interpretation. This interpretation assumes that the usable knowledge captured in teachers' individual rubric scores is different from the knowledge reflected in teacher's total clip scores. As a result, different dimensionality structures can be interpreted meaningfully because they reflect two different competencies. The multidimensional structure based on the individual rubric scores would be consistent with the understanding that teachers need to have different kinds of knowledge that they can apply to teaching situation. The single dimension underlying teachers' clip level scores, is assumed to capture teachers' ability to activate and strategically use their knowledge to produce interpretations that are most useful to understand a given teaching episode. This interpretation is consistent with a view on knowledge as a functional system, in which the meaning created by the analysis as a whole goes beyond the meaning conveyed by the individual knowledge components. Viewed from a system perspective, 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.

Classrooms are complex environments and for teachers to act in them effectively, simply having lots of knowledge is not going to be enough. In order to make informed instructional decisions, teachers need to be able to use their knowledge system efficiently to produce the most useful interpretation for a given context. If we assume that instructional decision making relies in part on the degree to which teachers' are able to make sense of teaching situations, then teachers' whole analyses provide a better basis for decision making than any or all of the individual knowledge components separately. At this point, we cannot say with certainty whether teachers clip level scores measure usable knowledge as a functional system. The factor analyses results by themselves do not provide sufficient evidence for this interpretation, but the two example responses we discussed show well, both that the entire analysis is more than the sum of its single components and that the usefulness and relevance of the analyses varies across teachers.

What is 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 of different levels of knowledge 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 the 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.

To be sure, the knowledge dimensions captured by the CVA scoring rubrics are certainly not the only dimensions that help describe the knowledge base from which teachers draw, and perhaps they are not the most important ones. We only score teachers responses according to four rubrics when in reality there are additional knowledge dimensions not captured by the CVA. Nevertheless, taking a functional system view on knowledge, however limited realized in the CVA, seems a promising avenue for understanding usable knowledge and how teachers use their knowledge in the classroom.

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

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

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 CT: Ablex, Ball, D. L. & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes and L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and pract*ice (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

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 and 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, NY: 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. Pfleiderer (Ed.) *New directions for teacher assessment* (Proceedings of the 1988 ETS Invitational Conference) (pp. 39–68). Princeton, NJ: 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

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.

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

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: 10.3102/00028312042002371

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.

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)

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., & 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

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

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. Stable URL: 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. & 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

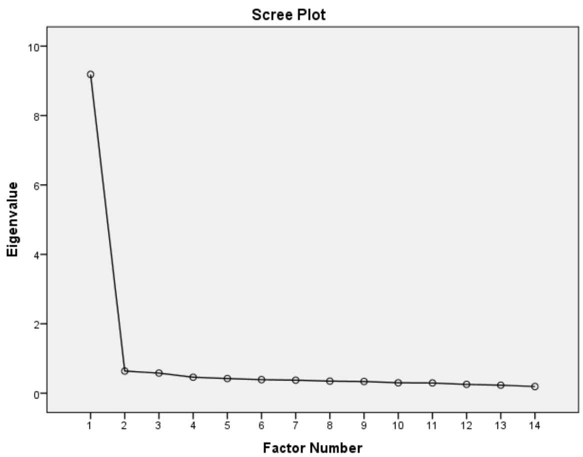
Stürmer, K., König, K., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *British Journal of Educational Psychology*, *83*(3), 467–483. doi: 10.1111/j.2044-8279.2012.02075.x van Es, E. A. & Sherin, M. G. (2008). Mathematics 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

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

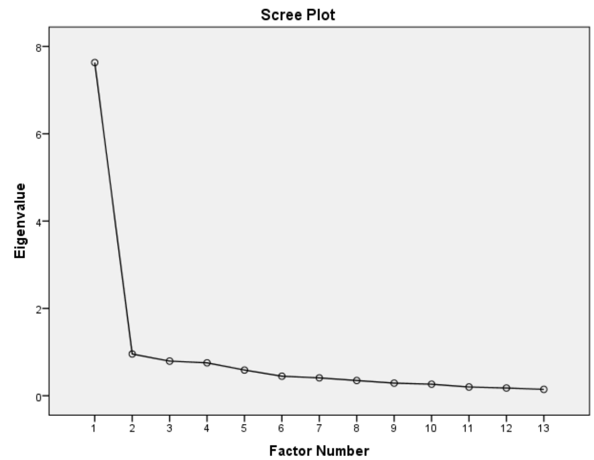
Appendix

Figures A1a-c.

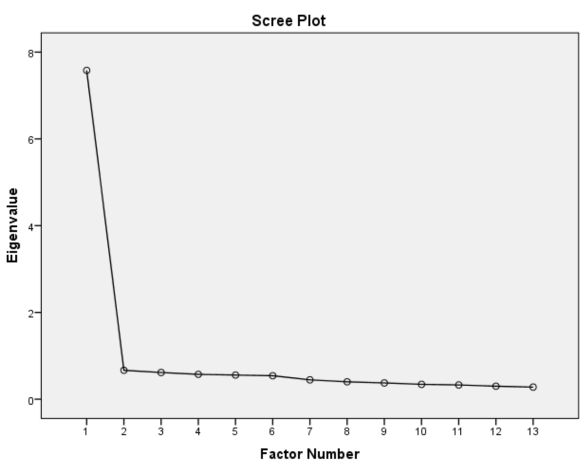
Scree Plots of the Extracted Factor Structure by Topic



1a. Variables, Expressions, and Equations



1b. Fractions



1c. Ratios and Proportions