Teaching and Teacher Education 63 (2017) 285–295

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Research paper

Prompting meaningful analysis from pre-service teachers using elementary mathematics video vignettes

Shereen Oca Beilstein [a], Michelle Perry [a,*], Meg Schleppenbach Bates [b]

[a] University of Illinois at Urbana-Champaign, 210 Education Building, 1310 S. Sixth Street, Champaign, IL, 61820, USA
[b] University of Chicago, CEMSE, 1427 E. 60th St., Chicago, IL, USA

HIGHLIGHTS

- Seeing short videos (<4 min) can produce analytical comments from pre-service teachers.
- Participants' responses appear to be malleable and sensitive to prompt types.
- Participants were most analytical when asked to focus on the teacher portrayed in the video.
- Participants were most descriptive when asked to focus on the students.

ARTICLE INFO

Article history:
Received 29 April 2016
Received in revised form
15 December 2016
Accepted 9 January 2017

Keywords:
Elementary mathematics
Video
Professional development

ABSTRACT

Learning from video is a theoretically grounded and popular professional development activity. In online professional development communities, however, responses to video are often shallow and lack meaningful commentary about issues that surround teaching and learning mathematics. By altering the framing conditions that accompany video clips posted to the *Everyday Mathematics* Virtual Learning Community, this study examined whether more deeply analytical comments could be elicited from pre-service teachers. Findings highlight the malleability of pre-service teachers' commentary, as their levels of analysis varied in relation to manipulations of perceived audience (expert or peer) and focus requested (on students, the teacher, or unspecified).

1. Introduction

Recruiting and developing highly qualified science, technology, engineering, and mathematics (STEM) teachers is an issue of national importance, as recent literature has demonstrated the need to improve the quality of instruction and the nature of interaction teachers provide their students, especially in elementary mathematics classrooms (Pianta, Belsky, Houts, Morrison, and the National Institute of Child Health and Human Development Early Child Care Research Network, 2007). Alongside the need to support STEM teachers, we note a recent rise of video-based learning increasingly becoming touted as one of the most highly effective practices for STEM teacher development. Importantly, research has linked changes in what teachers notice in classroom video to changes in their beliefs (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Chval, Lannin, Arbaugh, & Bowzer, 2009) and classroom practices (Sherin & van Es, 2009), which have been associated with improved student outcomes (Kersting, Givvin, Sotelo, & Stigler, 2010).

Video's potential to convey the “richness and complexities” of classroom interactions (Brophy, 2004, p. ix) has caused the medium to become an integral part of pre-service teacher education (Chval et al., 2009; Santagata & Angelici, 2010; Star & Strickland, 2008; Sun & van Es, 2015). But simply providing pre-service teachers with the opportunity to watch video does not automatically lead to their learning, and ultimately implementing, effective classroom practices. At the outset, pre-service teachers may not know what to focus on (Star & Strickland, 2008). Thus, learning how to notice key features of instruction and student thinking from video is an essential part of successful teacher preparation and education (Jacobs, Lamb, & Philipp, 2010; Santagata & Angelici, 2010; Sherin & van Es, 2009; Star & Strickland, 2008). Because the ability to analyze and learn from video clips is argued to be malleable,

* Corresponding author.
E-mail addresses: beilste2@illinois.edu (S.O. Beilstein), mperry@illinois.edu (M. Perry), megbates@uchicago.edu (M.S. Bates).

http://dx.doi.org/10.1016/j.tate.2017.01.005

researchers have examined practices designed to enhance this skill among pre-service teachers (Santagata & Angelici, 2010; Star & Strickland, 2008; Sun & van Es, 2015).

Although the bulk of literature surrounding the use of video in pre-service teacher education explores traditional, face-to-face settings, few studies have examined the use of video online—despite the number of high profile and popular online communities that allow teachers to interact with video (e.g., Inside Mathematics and the Teaching Channel). And despite significant investment in online communities, few studies have systematically investigated whether these resources are effective in promoting teacher learning (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009).

According to Ball and Cohen (1999), teachers often encounter difficulty critically analyzing specific elements of their own practice with peers—a possible by-product of the prevailing belief that "every teacher has to find his or her own style" (p. 19). Correspondingly, in online professional development communities, the commentary generated by teachers in response to video clips tends to be shallow, rarely engaging in the depth of analysis that leads to teacher learning (Schleppenbach & Beer, 2012; Kling & Courtright, 2003). This shortage of analytical responses to video clips in online communities does not necessarily indicate that learning is not occurring within the individual. At the individual level, teachers may be reacting to and learning from classroom interactions portrayed in video clips, but also may withhold from responding publicly online due to a number of reasons (e.g., fear of harming others' feelings, lack of time, etc.). Granted, it is possible that the analysis of video clips is not as common in online communities as professional developers and researchers in education would hope. Although research suggests that the analysis of video clips is a malleable skill (Jacobs et al., 2010; Santagata & Angelici, 2010; Sherin & van Es, 2009; Star & Strickland, 2008), its development through discourse requires guidance and facilitation (van Es, Tunney, Goldsmith, & Seago, 2014), which becomes an issue for online professional development sites where the presence of an expert facilitator or moderator is lacking (Bates, Phalen, & Moran, 2016). Indeed, one of the key questions about online learning from video is how can reflective commentary be produced in a setting where teachers view video asynchronously, with no guarantee of supportive facilitators or colleagues to push thinking forward.

Because videos posted to online professional development sites rarely produce meaningful commentary about the issues that surround teaching and learning mathematics, the impact of these clips on teacher learning is questionable. Thus, the goal of this study was to examine whether more deeply analytical comments could be elicited from pre-service teachers in response to video clips posted to the *Everyday Mathematics* Virtual Learning Community (VLC). Modeled after Hur and Hara's (2007) Korean online community, the VLC is a National Science Foundation-funded site with approximately 43,900 members. The present experiment was designed to address this overarching research question: Do differences in prompts accompanying video clips cause variations in pre-service teachers' analytical commentary? More specifically, by altering pre-service teachers' perceived audience and by shifting their attention to specific video elements, could more sophisticated commentary be compelled? By merging two avenues of research, video-based teacher education and online resources for professional development, we hope to contribute to extant literature (Borko, 2004; Borko et al., 2008; Brophy, 2004; Seago, 2004; Sun & van Es, 2015; van Es & Sherin, 2008) and to further understanding on how to promote analytical discourse and teacher learning online.

1.1. Video-based learning in the development of mathematics teachers

The use of classroom video clips in teacher development, or video-based learning, can be an effective tool to guide and refine those aspects of instruction that teachers notice—the goal being attending to and interpreting students' mathematical ideas (Jacobs et al., 2010; Santagata & Angelici, 2010; Sherin & van Es, 2009; Star & Strickland, 2008). Often used as case studies, video clips provide teachers with opportunities to analyze specific learning situations, consider the role and extent to which various classroom factors are involved, and consider alternate approaches and strategies to optimize student learning (Stigler & Perry, 2000; Brophy, 2004). In the past decade, and particularly in the field of mathematics education, video-based learning has become a prevalent area of research (Sherin, 2004) in both in-service teacher development (Jacobs et al., 2010; Sherin & van Es, 2009) and pre-service teacher education (Chval et al., 2009; Jacobs et al., 2010; Santagata & Angelici, 2010; Star & Strickland, 2008; Sun & van Es, 2015).

Here, we highlight some important recent findings. Sherin and van Es (2009) found that the dialogue among teachers in a video club progressed from initially focusing on instructional dimensions of the video—classroom management and environment, among others—and describing what had transpired to focusing on and attempting to understand the mathematical thinking displayed by students. Jacobs et al. (2010) captured, in a cross-sectional design, that differences in teachers' noticing and treatment of students' mathematical thinking were related to differences in their years of experience both in the classroom and with professional development. The more experience teachers had, Jacobs et al. (2010) concluded, the more likely they were to *attend*, *interpret*, and *decide how to respond* to relevant aspects of student thinking with specificity and skill. In addition, Santagata and Angelici (2010) demonstrated that the depth with which pre-service teachers, enrolled at an Italian public university, analyzed video increased when asked to consider the impact of instruction on student understanding and to suggest additional approaches for teaching. Further, Star and Strickland (2009) tracked changes in the classroom features pre-service teachers noticed and, subsequently, did not notice before and after they took a course designed to improve their observation skills. At the course's outset, pre-service teachers more readily focused on teacher actions related to administration and classroom management, whereas by the end, their noticing expanded to include other dimensions such as classroom environment (e.g., the layout of the classroom, class size, etc.), discourse, and subject matter. Central to each of these studies is the notion that the ability to focus on important or purposeful classroom activities is a malleable skill that can increase with experience and guidance.

Research also has indicated that improvements in how teachers view video clips are related to the use of effective classroom practices in mathematics (Sherin & van Es, 2009) and enhanced student outcomes (Kersting et al., 2010). Sherin and van Es (2009) applied Goodwin's (1994) framework—a discourse-based, socially situated process through which practitioners identify and analyze events relevant to their particular field—to teachers. By doing so, they developed a metric for teacher commentary on video, determining what teachers notice most and how teachers respond to what they notice. Sherin and van Es (2009) found that when teachers spent more time discussing student thinking in the professional development sessions, similar progress was observed in their own classroom instruction. For example, at the beginning of the study, one group of participants tended to superficially regard their students' ideas during classroom interactions. By the end, however, these teachers were more likely to respond to their students' ideas with interest, prompting further explanation and discussion. Expanding on these findings, Kersting et al. (2010) created a metric called Classroom Video Analysis that explored the relation between teachers' analyses of classroom video and student learning. Teachers' responses to the Classroom Video Analysis measure were positively correlated with their content knowledge for teaching and, importantly, were linked to student outcomes.

Much of the research in which associations have been made between improvements in teachers' analyses of video and shifts in their practice has focused primarily on in-service teachers. Sun and van Es (2015) recently extended these findings to pre-service teachers. In their study, pre-service teachers who learned to consider student thinking and to reflect on ways to improve student understanding during a video-based course were likely to implement practices that engaged their students, such as probing for student understanding, in their own classrooms.

1.2. Online resources for video-based teacher education and development

Past research supporting the notion that increased teacher reflection and pedagogical content knowledge are critical to promoting positive outcomes in the classroom has typically relied on intensive workshops—in traditional, face-to-face settings—as forums for professional development (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Fennema et al., 1996; Gearhart & Saxe, 2004). During the past decade, however, significant cuts to school-district budgets have rendered many workshops like these financially unfeasible. In light of the economic climate and in an effort to continue to promote professional development, online resources such as Inside Mathematics, the Teaching Channel, and the VLC have been developed, which provide teachers with free access to classroom video clips and other artifacts as they need it. Whereas traditional forms of professional development are temporal and static, these online resources are long term and continually accessible.

1.2.1. The challenges of building online communities

Although online resources have the potential to distribute video-based learning to a wider base of teachers than do traditional forums for professional development, less is known about the effectiveness of the online space in promoting teacher learning (Borko, Whitcomb, & Liston, 2009). Kling and Courtright (2003) studied the process by which an online group grows into a community by examining the Inquiry Learning Forum (ILF), a website with instructional materials, classroom video, and other resources for high school mathematics and science teachers. Of particular interest and concern was the discussion—the teachers' comments lacked the levels of engagement and depths of analysis for which site organizers had hoped. The interactions between teachers on the ILF were presided over by a sense of etiquette that inhibited analysis in their commentary (Kling & Courtright, 2003).

Additional research (Bates et al., 2016) has problematized the assumption that well-designed video-case studies automatically promote the same kind of learning on the web that they do in traditional settings. Bates et al. (2016) found that users of the VLC overwhelmingly responded to videos online by commenting on the teacher's pedagogy—rather than student thinking or mathematical content—and by evaluating the pedagogy, generally in a positive manner that is similar to "liking" a post or resource on the Internet. So although the VLC offers teachers access to learning from their peers, in essence creating a supportive community, this environment of support seems to have become a hindrance to effective teacher analysis, as the majority of commentary tends to be encouraging but not constructive (Schleppenbach & Beer, 2012). For example, in response to the "Animal Doubles" video clip posted to the VLC, wherein a first-grade mathematics teacher used a drawing of an animal to substantiate doubles facts, some comments were, "Using images like animals is a great idea for kids to become more excited and engaged in learning math," and, "Using animals, clever!" For this particular video clip, all of the responses from teachers were similar in that they commended the instructional practice demonstrated in the video, as opposed to considering its efficacy in promoting student understanding. These findings provide suggestive evidence that teachers may have difficulty learning from video clips in an online setting. At the very least, it appears that teachers are reluctant to provide comments that offer new insights about teaching or learning.

Part of the challenge encountered by teachers when learning from classroom video clips online could be accounted for by the absence of guidance and facilitation from knowledgeable teacher educators, who are central figures during traditional, face-to-face professional development programs (Bates et al., 2016). Take, as an example, the work by Sherin, van Es, and colleagues on video clubs, in which the benefits of high-level facilitation are evident (Sherin & van Es, 2009; van Es & Sherin, 2008; van Es et al., 2014). Often the club's facilitator would use an open-ended prompt (e.g., "What did you notice?") as a starter, a lead-in question to initiate discussion among group members (van Es & Sherin, 2008, p. 248). As the discussion progressed, the facilitator would then steer the dialogue toward analysis of what was noticed by weaving in questions that encouraged consideration of student thinking. Online, wherein the dialogue between commenters may not necessarily evolve over time, the utility of an open-ended prompt is questionable, especially for pre-service teachers, who may not yet know which aspects of classroom practice to attend to and interpret (Berliner, 1994; Jacobs et al., 2010; Santagata & Angelici, 2010; Star & Strickland, 2008). Indeed, developers of online communities must be resourceful when finding ways to promote meaningful analysis and reflection asynchronously. Although it is possible that community members may comment on a particular video, thus providing support for their colleagues' contributions, such interactions are not guaranteed and may happen over time periods too distant to help every teacher who views a video. Developers of online communities must find new structures that provide the kinds of support for video viewing generally given by facilitators or colleagues—by providing high-quality prompts and by developing features that mimic the commentary that would usually come from a facilitator or colleague. We thus follow up on this issue by systematically incorporating different prompts into our investigation.

1.2.2. Is there accountability online? The role of peers versus experts

According to social facilitation theory (Zajonc, 1965), the presence of an audience can induce changes in an individual's behavior; however, the mechanisms that elicit behavioral change are complex and varied, as performance may improve under some circumstances and decrease in others. And, according to Krauss (1987), the presence of an audience can shape the construction of individuals' messages in that the structure and substance of a message often will be influenced by commonalities they share with their perceived audience. The notion of social accountability then, defined by Tetlock (1983) as "pressures to justify one's opinion to others" (p. 74), could be interpreted as the merging of both Zajonc's and Krauss' ideas. Tetlock (1983) experimentally tested the construct of accountability by asking undergraduates to justify their opinions on controversial social issues by explaining them to someone who held liberal views, conservative views, unknown views, or under the condition of anonymity. In his study, participants' responses shifted in the direction of their perceived audience when required to justify their opinion to someone with a known viewpoint. Such theories of social facilitation and accountability could contribute to an explanation of the tenor of responses found from teachers on the VLC, as the assessment of one's peers can be tricky to navigate (Kaufman & Schunn, 2011). Members of the VLC may be inclined to construct their messages toward encouragement of peers and away from critique perhaps as a by-product of not wanting to offend or be unfair to their audience of fellow teachers. That is not to say that peer assessment and feedback do not have merits in the learning process; they have been regarded by participants as useful activities (Orsmond, Merry, & Reiling, 1996) that can affect the quality of pre-service teachers' learning and pedagogy (Sluijsmans, Brand-Gruwel, & van Merrienboer, 2002).

If we concede that the perceived audience of teachers' commentary on the VLC is composed of peers, then, hypothetically, what would happen to that commentary if the perceived audience were altered to include experts in teacher education? Theoretically, learning is enhanced when people know a knowledgeable audience has access to their explanations (Mero & Motowidlo, 1995; Rittle-Johnson, Saylor, & Swygert, 2008). To illustrate, undergraduates in Mero and Motowidlo's experiment (1995) were more accurate on a performance-rating task when they were told that they would have to explain their ratings to a more knowledgeable expert. Pushing these lines of inquiry further, it is plausible that the levels of analysis would increase among teachers on the VLC if they believed experts in education were going to read their commentary.

That we witness explicit analysis through commentary only rarely on the VLC does not necessarily mean analysis and learning are not occurring. It could be that teachers on the VLC are analyzing and learning from video clips, but are doing so implicitly, withholding their critiques of classroom practices or questions about student learning due to concerns that their comments may be received as unsupportive or rude. They may, in fact, be learning from the VLC, but not contributing to its community of learners.

1.3. The present study

Although extensive research has investigated the efficacy of face-to-face workshops using video to promote attentive noticing and meaningful analysis among teachers, few studies have considered this issue in the context of online professional development sites. To better understand how to promote analytical discourse in the online space, this study was undertaken to test whether it is possible to provoke teacher analysis through the use of prompts intentionally designed to shift their attention to various aspects of instruction. On the VLC and other online communities, the presence of peers potentially influences both the amount and degree of analytical commentary that members leave in response to video (Schleppenbach & Beer, 2012; Kling & Courtright, 2003). To this end, a central goal of this study was to determine whether teacher commentary could be shaped. To meet this goal, we gathered data from participants individually, with the intent of assessing what pre-service teachers observe and learn from video clips on the VLC, without peer interaction.

To address these research objectives, we experimentally manipulated prompts on two dimensions—perceived audience (expert or peer) and focus requested (on the student, on the teacher, or unspecified)—that have the potential to guide pre-service teacher commentary about classroom video clips.

Broadly, we anticipated that differences in perceived audience and foci requested would generate variations in participants' levels of analysis. More specifically, we generated three hypotheses based on theoretical perspectives and empirical findings emanating from research on teacher noticing and video analysis (e.g., Santagata & Angelici, 2010; Sherin & van Es, 2009; Star & Strickland, 2008; van Es & Sherin, 2008) as well as from research on social facilitation and accountability (e.g., Krauss, 1987; Mero & Motowidlo, 1995; Tetlock, 1983):

1. The open-ended prompt will result in less analytical responses than the teacher- or student-focused prompts.
2. Participants who are asked to direct their commentary toward experts in teacher education will provide more analysis than those who direct their commentary toward their peers, which is the current structure of the VLC.
3. An interaction between the peer-as-audience and unspecified prompt will lead to the least analytical commentary, while the expert-as-audience and student-focused prompt will compel the most analytical and sophisticated commentary.

2. Method

2.1. Participants

Ninety-four pre-service teachers enrolled in a College of Education degree-and-certification program at a large midwestern university participated. Approximately 97% of participants were female, and roughly 97% of participants were undergraduate students. The majority (69%) of participants were undergraduate seniors, and the average age of participants was 21.04 years. Participants were recruited from elementary mathematics methods courses and were offered extra credit in those courses in exchange for participation in the study. Instructors of these courses were asked to provide other forms of extra credit so as not to coerce student participation. The authors were not instructors in these courses.

2.2. Materials and procedure

Participants watched three video clips that were excerpted from first-grade mathematics classrooms and posted to the VLC (http://vlc.uchicago.edu). In each clip, the teacher used an example from real life to help students further understand and engage with a mathematical concept. All video clips lasted less than 4 min. Participants were asked to comment on the videos individually, outside of class. Participants watched the videos online from their own workplace—to mimic the structure of learning on the VLC.

2.2.1. On the selection of video clips

In face-to-face, video-based teacher development programs, much research has attended to the selection of clips (Borko et al., 2008; Brophy, 2004; Chval et al., 2009; Miller & Zhou, 2007; Seago, 2004), producing results that range from advocating for videos that depict best practice to those in which problems of practice arise. Videos on the VLC capture an array of teaching situations, including lessons that were successful and those that could be improved. The videos used in this study were of the latter sort, depicting a practice that is often utilized by teachers, but can be difficult to successfully implement: bringing mathematics into real-world contexts.

2.2.2. Perceived-audience and focus-requested manipulations

After viewing the videos, participants were asked to provide comments varying along two dimensions: perceived audience (peers or experts—across subjects) and focus requested (open ended, teacher focused, or student focused—within subjects).

Participants were randomly assigned to one of two perceived-audience conditions, peer or expert, and were given the following audience-framing prompt appropriate for that condition:

- *Peer as audience*: Many teachers find it useful to get feedback from other teachers. Please address your comments to other teachers, or
- *Expert as audience*: Many teachers find it useful to get feedback from experts in teacher training and education. Please address your comments to these experts.

Then, for each video, all participants were requested to focus on a particular aspect of the video by asking them to provide comments for one of the three foci:

- *Open-ended, or unspecified, prompt*: Comment on what you noticed about the examples in the video clip,
- *Teacher-focused prompt*: Comment on the teacher's use of examples to explain the math concept, or
- *Student-focused prompt*: Comment on the students' understanding of the examples to explain the math concept.

All participants watched the same three video clips and were asked to produce commentary for each of the three foci. Baseline data on participants' depth of commentary were not gathered beforehand in the form of a pretest.[1] We theorized that the peer-as-audience, open-ended prompt could serve as a (relative) baseline due to its lack of specificity, representing participants' analysis of video when unguided. Because the sequence in which the focus-related prompts were administered was not a central aim of this study, presentation of the video paired with a prompt was counterbalanced using a standard 3×3 Latin-squares design to help control for order (Keppel & Wickens, 2004). Within each square, video was counterbalanced so that, per condition, the three video clips were presented in one of three possible orders (ABC, BCA, CAB). Further, foci were counterbalanced to ensure that each prompt was distributed across different video clips and the sequence by which they were administered varied.

2.3. Coding

Each participant's response was coded for two dependent variables: central focus and level of analysis. Drawing on Sherin and van Es' (2009) *professional vision* metric, the coding scheme used in this study relied on two of their four dimensions: (1) the actor, which we refer to here as the central focus of the response, and (2) the commentary's stance, which we call level of analysis. First, the central focus of the response was coded either as teacher, student, or both. (It should be noted that Sherin and van Es' (2009) metric also included an additional actor code, other. However, in our data, so few responses were coded as other that we excluded this from our analyses.) Second, the level of analysis was coded either as description, evaluation, or interpretation. Descriptive responses reported on the actions and behaviors participants noticed in the video clips. Evaluative responses appraised the actions and behaviors participants noticed in the video clips. Interpretive responses included evaluations that either provided suggestions for improvement or made inferences based on evidence from the video clips. Although it may seem unusual to think about certain pairings of focus requested and their resulting levels of analysis (e.g., the student-focused prompt and interpretation), we treated this as an empirical question of whether each focus-requested prompt could actually elicit each level of analysis. Because previous research has treated these levels of analysis as ordered from low to high, description, evaluation, and interpretation (e.g., Bates et al., 2016; Palincsar, 1998; Sherin & van Es, 2009), responses in this study were coded for the highest level present. Table 1 provides examples of the coding scheme.

With 94 participants and three responses per participant, a total of 282 responses were produced. Responses were stripped of their prompts so that coders were unaware of which experimental condition they fell under, and two coders independently coded 22% of the full data set to establish inter-rater reliability. Substantial reliability was achieved on both dimensions: central focus (Cohen's kappa = 0.873) and level of analysis (Cohen's kappa = 0.912). After disagreements were resolved through discussion, the coders then went on to code the remaining data.

2.4. Manipulation fidelity: responsivity to focus requested

As a validity check, two tests were conducted to determine whether participants' responses were receptive to the focus-requested manipulation. First, the central focus of participants' responses was compared to the focus requested to determine their responsivity to this manipulation. In other words, when provided with a teacher-focused prompt, were participants referring to teacher behaviors in their responses? Or when provided with a student-focused prompt, were participants writing about student thinking? As the focus was unspecified for the open-ended prompt, those elements that the participants attended to were of major interest. Significant findings here indicated a dependent relation $[\chi^2(6) = 111.10,\ p < 0.001]$ between the focus-requested manipulation and the central focus of participants' responses (i.e., the teacher-focused prompt elicited teacher-focused responses and the student-focused prompt elicited student-focused responses). Further, the central focus of responses in relation to the open-ended prompt was mixed, either mainly focusing on the teacher only, student only, or evenly discussing aspects of both (see Fig. 1).

Next, to determine participants' responsivity to the open-ended prompt, which asked them to "Comment on what you noticed about the examples in the video clip," instances when participants used the word "notice" in their responses were counted. Because the open-ended request was the only prompting condition to use the word "notice," participants' subsequent use of the word "notice" in response to this prompt theoretically could gauge their receptiveness to this experimental manipulation. Participants used the word "notice" significantly more in relation to the open-ended prompt as opposed to the teacher- and student-focused prompts $[F(2) = 37.40,\ p < 0.001]$.

2.5. Data analysis

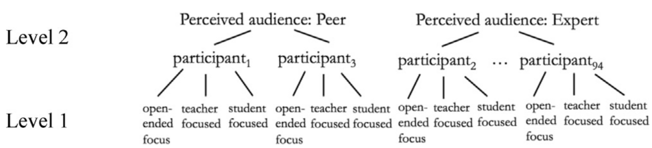
2.5.1. Statistical model

Our research questions utilized independent variables at two different levels: (1) *focus requested*, a within-subjects factor, with all three foci administered to all participants, and (2) *perceived audience*, a between-subjects factor randomly assigned across participants. Because of the design of this experiment, the manipulation of *focus requested* was nested within the manipulation of *perceived audience*, resulting in two hierarchical levels—the former qualifying as Level 1 and the latter as Level 2.

[1] We note that this was not a change-over-time study. Instead, it served more as a proof-of-concept study in which we investigated whether changes in prompts affect changes in teacher analysis.

Table 1
Examples of coded responses.

Response from participant	Central focus code	Level of analysis code
I noticed that when asked to create a "doubles" equation to be equivalent to the amount of eyes on the animal, students had difficulty deciding what math problem would match it.	Student	Description
The teacher's use of examples … to explain the math concept was very unique and effective.	Teacher	Evaluation
I think that the use of the dog is successful because kids can relate and understand math better when they see something that is familiar and visual… [The teacher] didn't draw the other eye, which made it harder to visualize… In order to be more successful, I think the teacher could incorporate different animals with high double numbers such as an octopus or spider.	Both	Interpretation



Furthermore, the experimental design produced a categorical dependent variable, *level of analysis*. Because the categorical dependent variable was not normally distributed (Snijders & Bosker, 2012) and because of the hierarchical nature of our manipulated independent variables, we used a hierarchical generalized linear model—more specifically, a multilevel multinomial logistic regression—in the analyses (Anderson, Kim, & Keller, 2013; Raudenbush & Bryk, 2002). The Level 1 independent variable of *focus requested* required participants to direct their commentary toward all three foci: the teacher, the student, and what they noticed in general. The Level 2 independent variable of *perceived audience* required participants to direct their commentary either toward their pre-service teacher peers or toward experts in teacher training and education. The dependent variable, *level of analysis*, included three types or values—description, evaluation, and interpretation—indicating the participants' depth of commentary in response to the three video clips viewed on the VLC.

2.5.2. *Modeling procedures*

Data were analyzed using the modeling procedures of multilevel multinomial logistic regression, as outlined by Raudenbush and Bryk (2002). The null model, Model 1, was examined first, assessing the outcome variable, level of analysis, while excluding all predictors at levels 1 and 2. Statistically significant variation for the perceived-audience manipulation was observed at this stage, indicating systematic variance between participants in the peer- and expert-as-audience conditions, thus warranting the use of multilevel analysis. Next, data were analyzed using only random intercepts by combining the null model with Level-1 fixed effects: These effects were related to the within-subjects, focus-requested manipulation (i.e., whether focus was on the teacher, on the student, or unspecified), resulting in Model 2. Results from this stage address Hypothesis 1, discussed earlier, which predicted that the open-ended prompt would lead to less analytical responses than the teacher- or student-focused prompts. Lastly, building on the prior two models, Level-2 fixed effects—those associated with the between-subjects, perceived-audience manipulation—were added, resulting in Model 3. Results from this stage address hypotheses 1 and 2, discussed earlier, describing the relations between the dimensions of perceived audience and focus requested and their resulting levels of analysis. Results also will indicate whether interaction effects exist within these independent variables, answering Hypothesis 3. Results from a comparison of Model 1, the null model; Model 2, which integrated the fixed effects from Level 1; and Model 3, which combined the fixed effects from levels 1 and 2, are reported in Table 3.[2] Based on the models' goodness-of-fit to the data—which included a log-likelihood ratio test that identified Model 3 as statistically different than Model 1—as well as their utility in addressing our hypotheses, Model 3 was the most appropriate to use when discussing the present study's results. For a more detailed account of the procedures used during model building and selection, please refer to the Appendix.

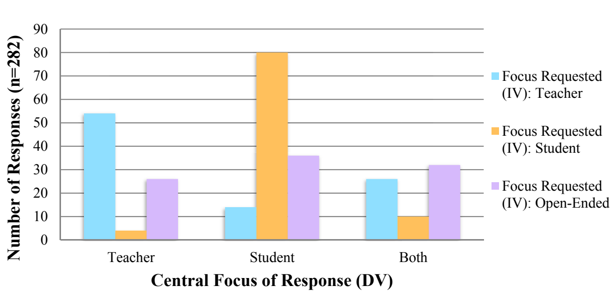


Fig. 1. Central focus of responses elicited by focus requested.

2.5.3. *Analysis software*

Proc GLIMMIX for generalized mixed models, in SAS 9.4, was used. Furthermore, maximum likelihood was estimated with the Gaussian quadrature as implemented in SAS 9.4 (Anderson et al., 2013).

3. Results

3.1. *Descriptive statistics*

The perceived-audience and focus-requested manipulations yielded frequencies for the dependent variable, level of analysis, found in Table 2. Note that each focus-requested manipulation yielded responses at each level of analysis.

3.2. *Effects of the focus-requested manipulation*

Model 3 was used to address Hypothesis 1, which predicted that the open-ended prompt would elicit fewer analytical responses than the teacher- or student-focused prompts in general, regardless of whether participants were randomly assigned to the peer- or expert-as-audience condition (see Fig. 2). Our findings supported this prediction as participants produced significantly more interpretive than descriptive responses when given a teacher-focused prompt as compared to an open-ended prompt [$t(90) = 3.31$, $p = 0.001$]. Stated in another way, the odds were 6.283 times more for participants to produce descriptive responses when their focus was open-ended as opposed to when

[2] All three models were computed using both robust and model-based estimation for standard errors. Differences in standard errors for the fixed effects across both types of estimation were minimal. The coefficients and standard errors reported in Table 3 were from model-based estimation.

Table 2
Frequencies for the dependent variable, level of analysis, by the independent variables: Perceived audience and focus requested.

Perceived audience	Focus requested	Level of analysis		
		Description	Evaluation	Interpretation
Peer	Open-ended	10	11	26
	Teacher	2	7	38
	Student	18	6	23
Expert	Open-ended	17	4	26
	Teacher	10	6	31
	Student	23	6	18

they were asked to focus on the teacher [95% CI = 2.088, 18.911].

Participants also produced significantly more interpretive than descriptive responses when given a teacher-focused prompt as compared to a student-focused prompt [$t(90) = 4.89,\ p < 0.001$]. In other words, the odds were 20.976 times more for participants to produce descriptive responses when their focus was on the students as opposed to the teacher [95% CI = 6.099, 72.144]. These results, then, did not support the hypothesis that the student-focused prompt would lead to the most sophisticated levels of analysis.

Two selections from two different participants demonstrate the variations in levels of analysis elicited by the focus-requested manipulation. Both selections were generated in response to the "Animal Doubles" video clip, wherein the teacher drew a dog on the chalkboard to help students understand the concept of doubles. When asked to focus on the teacher, one participant remarked,

> The teacher used an animal that the students are familiar seeing in their everyday life to find how they could use these "double facts" to explain the drawing. I think the use of "double facts" as a term was confusing to the students especially because there were parts of the animal that could have been doubles but were not seen easily, such as the eyes. Though this is only a short piece of a lesson I wonder if the teacher went on to use other animals as examples for double facts such as animals with only two legs, or insects with more than 2 eyes, and many legs. That I feel would have been a stronger way to show the connection with the doubles.

Coded as an interpretive response, this participant made inferences about the effectiveness of the teacher's instructional practice based on the evidence she observed from the video and also offered suggestions for improvement. When asked to focus on students' understanding, another participant observed,

> The students understood the first example 2 legs + 2 legs = 4 legs. However, when the teacher moved on to find other sets of things on the dog, like the eyes for example, the students had a hard time trying to stray away from the 2 + 2 = 4 example. When they heard the number 2, they wanted to do 2 + 2 = 4. However, once the teacher added the 2 + 2 to show 4 eyes and asked the students is this the double we want? 4 eyes? The students instantly understood what she was asking for and saw the double pair of 1 + 1 and also saw the connection to the 2 legs + 2 legs.

Coded as description, this participant reported on the actions she noticed in the video clip. One might argue that "students had a hard time" could suggest judgment or evaluation. This response, however, was coded as description because although there was a close and detailed observation of both the teacher's actions and student thinking displayed in the video, it lacked an interpretation of why this may have been the case.

3.3. Effects of the perceived-audience manipulation

Model 3 was also used to address Hypothesis 2, which predicted that participants in the expert-as-audience condition would produce more sophisticated commentary than those in the peer-as-audience condition. This finding was not statistically significant for description [$t(90) = 1.90,\ p = 0.060$] nor for evaluation [$t(90) = -0.64,\ p = 0.527$], which goes against our prediction that participants in the expert-as-audience condition would generate more sophisticated commentary than those in the peer-as-audience condition.

3.4. Interaction effects for the focus-requested and perceived-audience manipulations

No interaction effects between the perceived-audience manipulation, whether peer or expert, and the focus-requested manipulation, whether teacher focused, student focused, or open-ended, were present in relation to variations in levels of analysis. As a

Table 3
Parameter estimates for two-level generalized models of participants' levels of analysis ($n = 282$).

Parameter	Model 1	Model 2	Model 3[a]
Fixed effects	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Intercept 1 (Description)	−1.123** (0.28)	−3.232** (0.64)	−3.862** (0.78)
Intercept 2 (Evaluation)	−2.013** (0.39)	−2.361** (0.50)	−2.184** (0.54)
Level 1 (Focus requested)			
Description			
Open-ended		1.840** (0.55)	1.838** (0.55)
Student focused		3.048** (0.62)	3.043** (0.62)
Evaluation			
Open-ended		0.531 (0.48)	0.522 (0.48)
Student focused		0.525 (0.51)	0.509 (0.51)
Level 2 (Perceived audience)			
Description			
Expert			1.264 (0.67)
Evaluation			
Expert			−0.353 (0.56)
Error variance			
Level 2			
Subject & description	2.778** (1.20)	6.000* (2.54)	5.544* (2.39)
Subject & evaluation	2.265**(1.32)	2.468* (1.44)	2.404* (1.41)
Model Fit			
-2LL	507.75	469.71***	465.18
AIC	515.75	485.71	485.18

Note. Standard errors are in parentheses; Values based on SAS PROC Glimmix; Estimation method = Gaussian quadrature; Coefficients and standard errors calculated using model-based estimation.
* $p < 0.05$. ** $p < 0.01$. ***Likelihood ratio test significant.
[a] Best fitting model.

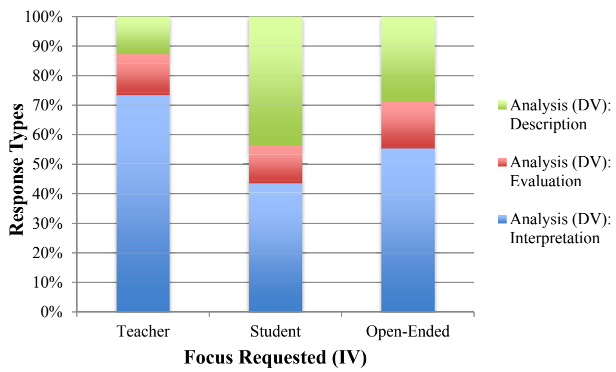


Fig. 2. Focus requested and levels of analysis.

result, we did not find support for an interaction between the perceived-audience and focus-requested prompts, as stated in Hypothesis 3.

4. Discussion and conclusions

The purpose of this study was to determine whether more deeply analytical commentary than what is naturally and typically produced on the VLC could be elicited from pre-service teachers individually. By altering the framing prompts that accompany video clips of mathematics classrooms, this study tested whether participants' levels of analysis varied in response to manipulations of perceived audience and focus requested. Three hypotheses were proposed. First, we predicted that the open-ended prompt would generate the least sophisticated commentary from participants. Second, we predicted that participants in the expert-as-audience condition would produce more sophisticated commentary than those in the peer-as-audience condition. Third, we predicted an interaction to occur between the perceived-audience and focus-requested manipulations, wherein the peer-as-audience and open-ended prompt would lead to the least analytical commentary, while the expert-as-audience and student-focused prompt would compel the most analytical and sophisticated commentary.

Overall, our findings indicate that viewing short classroom video clips can produce analytical commentary from pre-service teachers. Addressing Hypothesis 1, participants' responses appear to be malleable, as the prompt asking them to focus on the teacher elicited higher levels of analysis than prompts asking for other foci. It is possible that because participants were pre-service teachers, they were highly attuned to examining teacher behavior and pedagogy, thus more likely to make inferences about the teacher moves they viewed in the video clips than about student learning. We also anticipated that the student-focused prompt would yield more sophisticated commentary than the open-ended prompt, as research on video-based learning in professional development has associated increased consideration of student thinking and understanding with increased levels of teacher reflection (e.g., Santagata & Angelici, 2010; Sherin & van Es, 2009; van Es & Sherin, 2008). Yet contrary to our expectations, participants produced the most descriptive commentary when asked to focus on the students' understanding as depicted in the video, similar to Bates et al. (2016) findings. It must be noted, however, that these descriptions were not simple, as implied in other reports (e.g., Bates et al., 2016; Sherin & van Es, 2009), but rich. And although we followed previous research in organizing our coding from description to evaluation to interpretation, our results suggest that these are different types of reactions to classroom video, but are not necessarily ordered from lowest to highest. Detailed and thoughtful description may be a fundamental precursor to interpretation of student thinking and may be an appropriate level of analysis for novice or pre-service teachers.

Results pertaining to Hypothesis 2 were most surprising, given the large body of work on social facilitation and accountability, which explores how the presence of an audience can alter individuals' behaviors (e.g., Krauss, 1987; Mero & Motowidlo, 1995; Rittle-Johnson et al., 2008; Tetlock, 1983; Zajonc, 1965). We did not find significant differences in the levels of analysis between participants in the peer- or expert-as-audience conditions. It is feasible that these perceived-audience manipulations were not authentic. Pre-service teachers may have anticipated that those conducting the study were experts and would read their entries—therefore rendering the peer-as-audience manipulation ineffective. Or if the perceived-audience manipulation were effective, then data would suggest that the perceived presence of experts inhibited performance: Participants may have been intimidated by their presumed evaluators and, as a result, been less engaged (see Geen, 1991, for a review on social facilitation and social anxiety) and less inclined to contribute more analytical commentary. Although plausible, that Hypothesis 2 was not supported speaks to the complexity of accountability as a construct (Lerner & Tetlock, 1999). Unfortunately, fidelity assessment of the audience manipulation was missing in this experiment. Future research should further investigate the roles that anticipating peers' and experts' access to commentary play in shaping teachers' responses to video clips both when responding offline and on online professional development sites.

Hypothesis 3 was not supported by the data because the interaction effects were not significantly present in our analyses.

This study had several limitations. One limitation stems from the sample, which was limited to pre-service teachers. Because the membership base of the VLC consists of both pre- and in-service teachers, future research should examine the effects of different prompting conditions on in-service teachers as well. An additional shortcoming of the experiment was the temporality of its design. Future research would benefit from investigating whether the effects of participation on the VLC and their relation to teacher learning are sustained over time.

We also note that the focus-requested prompts were general by design—they did not probe teachers to consider specific points related to a particular video. Our intent was to have these prompts serve as a baseline from which the impact of other variables, such as prompt specificity, could be explored in future studies. For example, perhaps by increasing the prompts' specificity (e.g., "What misconceptions does Student X have?" Or, "What additional strategies might this teacher use to help further students' understanding?"), we can more readily incorporate those successful moves that facilitators make in person to the online space (e.g., Sherin & van Es, 2009; van Es et al., 2014).

We recognize that exploring the connection between teacher commentary on the VLC and student learning is a critical next step—one that could be examined from multiple angles. Future research should not only measure the correlation between the level of teacher analysis in response to videos on the VLC with student achievement (see Kersting et al., 2010), but also uncover the mechanisms by which teachers apply what they learn from the VLC to their own instruction—and more specifically, how that relates to the classroom environment (Sherin & van Es, 2009; Sun & van Es, 2015), student engagement, and achievement. Future research also should determine the impact of analytical commentary on (a) respondents' pedagogical content knowledge; (b) others' (e.g., other teachers who read the commentary on the VLC) pedagogical content knowledge; and (c) feelings of belonging (or not) to a community interested in teaching and learning mathematics.

In this investigation, participants provided their comments in isolation—data were gathered from participants individually and offline, apart from the VLC. Furthermore, participants were aware that their comments would not be uploaded to the VLC. By designing the experiment in this manner, our intent was to determine whether teacher commentary could be shaped through the manipulation of two dimensions: perceived audience and focus requested. Although we were successful in this limited context, other dimensions such as the role facilitators and peers play in shaping discourse patterns as well as the interaction that results from posting comments publicly to the VLC are other key dimensions that need to be explored.

Online professional development sites increasingly are becoming popular resources for teachers. On the VLC, teachers from across the globe can access classroom video clips and interact with their peers—not only locally, but also internationally. These sites, however, are an under-researched area of teacher education (Dede et al., 2009). By promoting analytical and insightful teacher commentary in the online space, teachers, in turn, may be prompted to reflect on and analyze their own pedagogy, and further refine their classroom practices. And, by turning their attention to the cognitive processes and abilities of students during instruction (Ball, Lubienski, & Mewborn, 2001; Carpenter et al., 1989; Fennema et al., 1996; Gearhart & Saxe, 2004), teachers can become more adept at anticipating and addressing the potential challenges in student understanding within a specific lesson (Gearhart & Saxe, 2004). When teachers learn, not only through analysis of their own practices, but also from their peers, the benefits for the classroom can be substantial. This study aimed to uncover the particular contexts that potentially could provoke insightful commentary, which could be posted and viewed by participating teachers on a VLC. This investigation thus represents a fundamental step toward better understanding the role and magnitude of sites like the VLC in teacher education and online professional development.

Acknowledgments

This work was supported in part by the University of Illinois Urbana-Champaign Campus Research Board, Grant No. RB15002 and the National Science Foundation, Grant No. 1621253. We owe gratitude to the pre-service teachers who took time to view the videos and provide comments about what they saw, which constituted the data source for this research, as well as all those who worked on the VLC's development at UChicago STEM Education. We also thank Carolyn Anderson, R. Wes Crues, and Shuai Wang for sharing their methodological expertise; Juan Gerardo, Martha Makowski, Adam Poetzel, Stephanie Saclarides, and Valerie Willetts for their assistance in data collection; and Neet Priya Bajwa, Paul Beilstein, Jeanne Brunner, Gloriana González, Genevieve Henricks, Sarah Lubienski, Martha Makowski, Joseph Robinson-Cimpian, and Shuai Wang for thoughtful feedback provided at various points throughout the research process.

Appendix

Modeling procedures

The null model

Because the dependent variable was polytomous, consisting of three levels of analysis (e.g., description, evaluation, and interpretation), two estimates were needed when evaluating the null model. Both estimates—the log odds of producing descriptive versus inter-pretive commentary and the log odds of producing evaluative versus interpretive commentary—were calculated simultaneously and in relation to the referent category (K), which was interpretation. Equation 1 through 5 represent the models used during this stage.

Model 1: The null model.

Level-1 link function:

$$logit(Y_{ij} = k) = \eta_{kij} = log\left(\frac{P(Y_{ij} = k)}{P(Y_{ij} = K)}\right) \tag{1}$$

where

- The outcome at Level 1 is the log odds of category *k*, description or evaluation, relative to the referent category *K*, interpretation

Level-1 cluster-specific model:

$$P(Y_{ij} = k) = \frac{\exp\left(\eta_{kij}\right)}{\sum_{k=1}^{K} exp\left(\eta_{kij}\right)} \tag{2}$$

Level 1:

$$\eta_{kij} = \beta_{0j(k)} \tag{3}$$

Level-2 model:

$$\beta_{0j(k)} = \gamma_{00(k)} + u_{0j(k)} \tag{4}$$

where

$$\left(u_{0j(k)}\right) \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\tau_{00(1)00(1)} & \tau_{00(1)00(2)}\\ \tau_{00(2)00(1)} & \tau_{00(2)00(2)}\end{pmatrix}\right]$$

- $\beta_{0j(k)}$ = model intercept for participant *j*, representing the log odds for the comment's level of analysis *k* (whether description vs. interpretation or evaluation vs. interpretation)
- $\gamma_{00(k)}$ = mean of intercepts across both perceived-audience conditions
- $u_{0j(k)}$ = random variation of intercepts across both perceived-audience conditions

Mixed linear predictor model:

$$log\left(\frac{P(Y_{ij} = k)}{P(Y_{ij} = K)}\right) = \gamma_{00(k)} + u_{0j(k)} \tag{5}$$

Results from the null model indicated that the log odds of producing descriptive commentary were less than the log odds of producing interpretive commentary $[\hat{\gamma}_{00(1)} = -1.123,\ t(186) = -4.03,\ p < 0.001]$. Additionally, the log odds of producing evaluative commentary were less than the log odds of producing interpretative commentary $[\hat{\gamma}_{00(2)} = -2.013,\ t(186) = -5.16,\ p < 0.001]$. There appears to be statistically significant variation between the perceived-audience conditions in the log odds of producing descriptive comments relative to interpretive ones $[\hat{\tau}_{00(1)00(1)} = 2.777,\ z = 2.31,\ p_{mixture\ from\ \chi^2_1\ and\ \chi^2_2} < 0.001]$ (Snijders & Bosker, 2012). In addition, there appears to be statistically significant variation between the perceived-audience conditions in the log odds of producing evaluative comments relative to interpretive ones $[\hat{\tau}_{00(2)00(2)} = 2.265,\ z = 1.71,\ p_{mixture\ from\ \chi^2_1\ and\ \chi^2_2} < 0.001]$.

The conditional model

Model 2. Model 2 built on the null model by adding the Level-1 mixed effects of the focus-requested manipulation (i.e., whether focus was on the teacher, on the student, or open-ended).Equations (6) through (9) depict the models utilized during this step.

Level-1 model with Level-1 predictors only [Model 2]

$$log\left(\frac{P(Y_{ij}=k)}{P(Y_{ij}=K)}\right)=\eta_{kij}=\beta_{0j(k)}+\beta_{1j}(Focus\ Requested)_{ij} \quad (6)$$

where

- $\beta_{1j}=$ regression coefficient representing the effect of the within-subject independent variable of focus requested (teacher focused, student focused, or open-ended) for the comment's level of analysis k

Level-2 model:

$$\beta_{0j(k)}=\gamma_{00(k)}+u_{0j(k)} \quad (7)$$

$$\beta_{1j}=\gamma_{10} \quad (8)$$

where

$$\left(u_{0j(k)}\right)\sim N\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\tau_{00(1)00(1)} & \tau_{00(1)00(2)}\\ \tau_{00(2)00(1)} & \tau_{00(2)00(2)}\end{pmatrix}\right]$$

- $\beta_{0j(k)}=$ model intercept for participant j, representing the log odds for the comment's level of analysis k (whether description vs. interpretation or evaluation vs. interpretation)
- $\beta_{1j}=$ regression coefficient representing the effect of the within-subject independent variable of focus requested (teacher focused, student focused, or open-ended) for the comment's level of analysis k
- $\gamma_{00(k)}=$ mean of intercepts across both perceived-audience conditions
- $\gamma_{10}=$ regression coefficient of the perceived-audience manipulation
- $u_{0j(k)}=$ random variation of intercepts across both perceived-audience conditions

Mixed linear predictor model:

$$log\left(\frac{P(Y_{ij}=k)}{P(Y_{ij}=K)}\right)=\gamma_{00(k)}+\gamma_{10}(Focus\ Requested)_{ij}+u_{0j(k)} \quad (9)$$

According to Equation (6), the intercept $\beta_{0j(k)}$ and slope (regression coefficient β_{1j}) are dependent on condition k, which was participants' level of analysis, either description or evaluation. During this step of model building, these coefficients can vary across between-subject conditions, allowing for the analysis of Level-1 fixed effects only (Raudenbush & Bryk, 2002).

Model 3. Model 3 added the Level-2 fixed effects associated with the perceived-audience manipulation. Here, the Level-1 model as depicted in Equation (6) remains the same. The Level-2 model, however, is expanded as represented in equations (10) through (12).

Level-2 model:

$$\beta_{0j(k)}=\gamma_{00(k)}+\gamma_{01}(Perceived\ Audience)_{j}+u_{0j(k)} \quad (10)$$

$$\beta_{1j}=\gamma_{10} \quad (11)$$

where

$$\left(u_{0j(k)}\right)\sim N\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\tau_{00(1)00(1)} & \tau_{00(1)00(2)}\\ \tau_{00(2)00(1)} & \tau_{00(2)00(2)}\end{pmatrix}\right]$$

Mixed linear predictor model:

$$log\left(\frac{P(Y_{ij}=k)}{P(Y_{ij}=K)}\right)=\gamma_{00(k)}+\gamma_{10}(Focus\ Requested)_{ij}+\gamma_{01}(Perceived\ Audience)_{j}+u_{0j(k)} \quad (12)$$

For Model 3, there appears to be statistically significant variation between participants in the log odds of producing a descriptive comment relative to an interpretive one [$\hat{\tau}_{00(1)00(1)}=5.544$, $z=2.32$, $p_{mixture\ from\ \chi^2_1\ and\ \chi^2_2}<0.001$]. In addition, there appears to be statistically significant variation between participants in the log odds of producing an evaluative comment relative to an interpretive one [$\hat{\tau}_{00(2)00(2)}=2.404$, $z=1.70$, $p_{mixture\ from\ \chi^2_1\ and\ \chi^2_2}<0.001$].

On model development and the selection of the final model

To determine whether random effects were needed, the significance of the random intercept variance parameters were tested for each model, as seen above, by mixing the χ^2_p and χ^2_{p+1} distributions (where p refers to the number of covariances) and then taking the average of each distribution's p-values (Snijders & Bosker, 2012).

To determine goodness of fit among the three models, deviance tests comparing differences among −2 log likelihood ratios (-2LL) between nested models—Model 1 versus Model 2 and Model 2 versus Model 3—were conducted using the following equation:

$$\chi^2_{diff}=-2\ Log\ Likelihood_{(Null)}—2\ Log\ Likelihood_{(Full)}$$

The -2LLs for models 1, 2, and 3 were 507.75, 469.71, and 465.18, respectively. Differences between the -2LLs produced by the nested models indicated that Model 2, the Level-1 fixed effects model, produced a significantly better fit than Model 1, the null model [$\chi^2(4)=38.04$, $p<0.001$], but the difference between Model 2 and Model 3, which combined levels 1 and 2, was statistically insignificant [$\chi^2(2)=4.53$, $p=0.104$]. Although this test would suggest that Model 2 should be our final model, because Model 3 had the smallest Akaike information criterion and because it allowed us to answer the hypotheses proposed as part of this study, Model 3 was deemed the most appropriate to use when deriving statistical interpretations of the data.

References

Anderson, C. A., Kim, J., & Keller, B. (2013). Multilevel modeling of categorical response variables. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 481–519). Boca Raton, FL: Taylor and Francis.

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., Lubienski, S., & Mewborn, D. (2001). Mathematics. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). Washington, DC: American Educational Research Association.

Bates, M. S., Phalen, L., & Moran, C. (2016). If you build it, will they reflect? Examining teachers' use of an online video-based learning website. *Teaching and Teacher Education, 58*, 17–27.

Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. In J. M. Mangier, & C. C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161–186). Fort Worth, TX: Harcourt.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15.
Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education, 24*(2), 417–436.
Borko, H., Whitcomb, J., & Liston, D. (2009). Wicked problems and other thoughts on issues of technology and teacher learning. *Journal of Teacher Education, 60*(1), 3–7.
Brophy, J. (2004). Advances in research on teaching. In *Using video in teacher education* (Vol 10). Oxford, UK: Elsevier.
Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499–531.
Chval, K. B., Lannin, J. K., Arbaugh, F., & Bowzer, A. D. (2009). Videos and prospective teachers. *Teaching Children Mathematics, 16*(2), 98–105.
Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education, 60*, 8–19.
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*, 244–276.
van Es, E. A., Tunney, J., Goldsmith, L., & Seago, N. (2014). A framework for the facilitation of teachers' analysis of video. *Journal of Teacher Education, 65*, 340–356.
Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403–434.
Gearhart, M., & Saxe, G. B. (2004). When teachers know what students know: Integrating mathematics assessment. *Theory into Practice, 43*, 304–313.
Geen, R. G. (1991). Social motivation. *Annual Review of Psychology, 42*, 377–399.
Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*, 606–633.
Hur, J. W., & Hara, N. (2007). Factors cultivating sustainable online communities for K-12 teacher professional development. *Journal of Educational Computing Research, 36*, 245–268.
Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*, 169–202.
Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science: An International Journal of the Learning Sciences, 39*, 387–406.
Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Pearson.
Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teachers' analysis of classroom video predicts student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*, 172–181.
Kling, R., & Courtright, C. (2003). Group behavior and learning in electronic forums: A sociotechnical approach. *The Information Society, 19*, 221–235.
Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology, 6*, 81–98.
Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255–275.
Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology, 80*, 517–524.
Miller, K. F., & Zhou, X. (2007). Learning from classroom video: What makes it compelling and what makes it hard. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 321–334). Mahwah, NJ: Erlbaum.
Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*, 239–250.
Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49*, 345–375.
Pianta, R. C., Belsky, J., Houts, R., Morrison, F., & The National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*, 1795–1796.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Rittle-Johnson, B., Saylor, M., & Swygert, K. E. (2008). Learning from explaining: Does it matter if mom is listening? *Journal of Experimental Child Psychology, 100*, 215–224.
Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on pre-service teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education, 61*, 339–349.
Schleppenbach, M., & Beer, D. (2012). Designing a virtual learning community for elementary mathematics teachers. In *Paper presented at the American Educational Research Association annual meeting, Vancouver, British Columbia*.
Seago, N. (2004). Using video as an object of inquiry for mathematics teaching and learning. In J. Brophy (Ed.), *Vol 10. Advances in research on teaching* (pp. 259–286). Oxford, UK: Elsevier. Using video in teacher education.
Sherin, M. G. (2004). New perspectives on the role of video in teacher education. In J. Brophy (Ed.), *Using video in teacher education: Vol. 10. Advances in research on teaching* (pp. 1–27). Oxford, UK: Elsevier.
Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*, 20–37.
Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merrienboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education, 27*, 443–454.
Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.
Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve pre-service mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*, 107–125.
Stigler, J. W., & Perry, M. (2000). Developing classroom process data for the improvement of teaching. In N. S. Raju, J. W. Pellegrino, M. W. Bertenthal, K. Mitchell, & L. Jones (Eds.), *Grading the Nation's Report Card: Research from the evaluation of NAEP* (pp. 229–264). Washington, DC: National Academy Press.
Sun, J., & van Es, E. A. (2015). An exploratory study of the influence that analyzing teaching has on preservice teachers' classroom practice. *Journal of Teacher Education, 66*, 201–214.
Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*, 74–83.
Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274.