

An Analysis of Evaluative Comments in Teachers' Online Discussions of Representations of Practice

Vu Minh Chieu

Karl W. Kosko

Patricio Herbst

University of Michigan

Kent State University

University of Michigan

Abstract: It has been common to use video records of instruction in teacher professional development, but participants have rarely been encouraged to evaluate teachers and students' actions in those records, allegedly because evaluation deters from the development of a professional discourse. In this study, we inspected teachers' online discussions of animations of classroom episodes realized with cartoon characters, looking at the difference in the content of conversation turns when members made evaluative comments and when they did not make evaluative comments. We were interested in finding out whether making evaluative comments correlated with participants' reflection on their professional practice and proposal of alternative teaching actions; for that purpose we used systemic functional linguistics to develop a coding scheme that attended to evaluation, alternatives, and reflection in forum discussions. We found statistically significant evidence that the more the participants actively evaluated the teaching in the animations the more they proposed alternative teaching actions and reflected on instructional practice. We relate these findings to the notion of social presence in online discussions.

Keywords: asynchronous discussion, multilevel models, systemic functional linguistics, teacher education, evaluation, animations, social presence.

Introduction

Both online and offline discussions have been commonly used for teacher learners to exchange ideas and learn from one another about their professional practice (Barab, Kling, & Gray, 2004; Fishman & Davis, 2006). Shared resources (e.g., specific artifacts that illustrate or elicit professional knowledge) are instrumental to make members' conversations meaningful by inviting them to express personal ideas, to ask important questions, to comment on others' ideas, and so forth (Wise, Padmanabhan, & Duffy, 2009).

Video records of classroom interaction have often been used, as shared resources, to create face-to-face conversation contexts, for both preservice and in-service teachers to develop and elicit their professional knowledge and skills (e.g., Nachlieli & Herbst, 2009; Star & Strickland, 2008; van Es & Sherin, 2008). Also, researchers and teacher developers have exploited information and communication technologies to sustain teachers' online discussions, hence helping them to exchange, share, and learn about practical knowledge and skills with each other and with educational researchers and teacher educators (Barab, et al., 2004; Fishman & Davis, 2006).

Although video records of practice have been and continue to be useful in supporting teachers' discussions about their professional practice, members of discussion groups have rarely been encouraged to evaluate teachers and students' actions in this kind of representation of practice, mainly because of the perception that such evaluations might deter from the development of a professional discourse (Jacobs, Borko, & Koellner, 2009; Seago, 2004). So, researchers of video-based professional development have had little chance to examine the role of evaluation in teachers' discussions. Recently, researchers have started to use animations of classroom scenarios in sustaining teachers' conversations about instructional practice (e.g., Herbst & Chazan, 2006; Herbst & Miyakawa, 2008; Herbst, Nachlieli, & Chazan, 2011; Moore-Russo & Viglietti, 2011; Moore-Russo & Wilsey, 2014; Tettegah, 2005). In our studies with animated classroom stories (Chieu, Herbst, & Weiss, 2011; Chieu, Aaron, & Herbst, 2013; Chieu & Herbst, 2013), we have observed that participants frequently made evaluations of the actions of the cartoon teacher. More importantly, we have noticed that making such evaluations can be positive in that it may go along with participants' reflection on teaching actions or discussion of possible alternatives in teaching. This paper investigates the role of evaluation in the postings made in a set of eight online forums associated with animated classroom stories, by comparing the quality of those postings where participants made evaluative comments with those postings where they did not make evaluative comments. Two of the desirable features of teacher discourse identified by Herbst and Chazan (2006), reflection and alternativity, were operationalized to account for the quality of postings. We examined the correlations between *evaluation*, as an indicator, and the probabilistic presence of *reflection* (whether or not the participants reflect on teaching actions that they notice in the classroom stories embedded into their discussion space) and *alternativity* (whether or not they propose alternative actions of

teaching when they discuss a teaching decision in the embedded animations), as two features in professional discourse about practice.

Theoretical Framework and Related Work

Many kinds of technologies and communities have been implemented to support groups of both teacher candidates and practicing teachers in learning to do or in improving how they do the work of teaching (Barab, et al., 2004; Fishman & Davis, 2006). In this section, we narrow our review to the use of technologies for sustaining collaborative learning by teachers. We especially look at how other researchers in the field study qualities of discussions in teacher learning, with a particular attention to correlations between evaluative comments and those qualities.

Video technologies have been a common choice for teacher educators to support teachers' reflection, and their learning to notice and interpret critical areas of classroom interaction (Rich & Hannafin, 2009; Sherin, Jacobs, & Philipp, 2011). Teachers have been often organized into face-to-face groups, with one or two facilitators, to view, examine, and discuss video records of teaching (e.g., Herbst & Chazan, 2003; Nachlieli, 2011; Star & Strickland, 2008; van Es & Sherin, 2008; Zhang et al., 2011). Those artifacts have sometimes included video captures of their own teaching, records of their peers' instructional practice, or purportedly exemplary video records of teaching provided by third-party organizations (Zhang et al., 2011). An important characteristic of video records of classroom interaction is that they provide support for teachers to examine tactical and temporal entailments of instructional practice. The possibility to replay videos can help viewers spot important moments of a teaching episode and examine and discuss teaching tactics and strategies, student thinking, and so forth (Lampert & Ball, 1998).

An important assumption beneath the practice of using media representations of teaching to provoke discussions about teaching and eventually increase teaching capacity is that individuals who interact with this media can engross themselves with the actual practice being represented. Communication theorists have developed a variety of conceptions of *presence* (including *social presence*) to operationalize different ways in which individuals live the illusion that a mediated experience is not mediated (Lombard & Ditton, 1997; Oztok & Brett, 2011). In the case of teachers interacting with a representation of teaching and with each other, it is important to identify dimensions of those interactions that might contribute to such a sense of presence. Herbst & Chazan (2006) proposed a number of those dimensions, including *alternativity* (i.e., the capacity to consider alternative actions of teaching) and *reflection* (i.e., the capacity to make professional judgments of actions of teaching), as important qualities to look for in teachers' discussions of representations of teaching, qualities that we argue to be indicators of social presence.

Many studies in the literature on teacher learning have acknowledged the importance of supporting participants' reflection (Rich & Hannafin, 2009; Sherin et al., 2010; Schön, 1983; Zhang et al., 2011) and alternativity (Nespor, 1987; Scott, 2005; Wilkins, 2008). Yet, evaluation has not been considered systematically. Scholars do not seem to agree on the value of engaging viewers of classroom videos in evaluating features of classroom interactions. On the one hand, the video club study (van Es & Sherin, 2008) has included evaluation as an outcome variable in its coding scheme. Also, Males, Otten, and Herbel-Eisenmann (2010) used a critical collegueship framework in a way that promotes critical reflection and avoids damaging personal relationships within a group of mathematics teachers. On the other hand, in some teacher development environments where teachers view each other's video records, facilitators often

discourage participants from evaluating others' teaching, allegedly to promote sensitivity and professional discussion (Jacobs, Borko, & Koellner, 2009). For example, LeFevre (2004) has emphasized “the importance of relational aspects” (p. 254) in the facilitation of discussions about classroom videos of unknown teachers and students and noted the importance to enable “teachers to be critical about teaching in a non-judgmental manner” (p. 254). Similarly, Barnett (1987) and Joyce and Showers (1980) have encouraged non-evaluative feedback among peers in coaching of leadership and teaching. Knight (2006, p. 36) has also added: “Coaching is a non-evaluative, learning relationship between a professional developer and a teacher, both of whom share the expressed goal of learning together, thereby improving instruction and student achievement.” Another study (Kelchtermans & Vandenberghe, 1994) has collected data about field notes of classroom teaching and interviews with individual teachers to create and give feedback on a professional development story for each teacher. A rule of thumb for the researcher who commented on the story was to avoid judgments and to adopt a non-evaluative attitude. In addition, in teachers’ discussion about their professional practice even if the facilitator does not explicitly discourage it, participants have been reluctant to evaluate human teachers' practice (Seago, 2004; Zhang et al., 2011).

The tendency of avoiding judgments and adopting a non-evaluative attitude has existed not only in face-to-face conversations but also in online discussions. For instance, studies of web-mediated consultation conditions (Hadden & Pianta, 2006; Pianta et al., 2008) in teacher development have made a recommendation of “establishing a non-judgmental and non-evaluative supportive relationship” between a teacher and a consultant. Stiler and Philleo (2003) transformed strategies in face-to-face discussions to the design of a reflective, non-evaluative environment that supported candidate teachers in creating web-based journals or blogs about the

practice of teaching. To foster communication and reflective inquiry about different perspectives on the realities of classroom practice, Edens (2000) engaged teacher education students in a non-evaluative setting where they could share their observations and concerns comfortably. Colasante (2011) also promoted a non-judgmental or “safe” environment where preservice teachers in physical and sport education used an online media annotation tool to reflect upon their teaching practice (in the form of video records) individually or collectively.

Although the above literature review has indicated a number of supporting arguments and evidence for promoting non-evaluative¹ comments in both face-to-face and online conversations about the work of teaching, there are important reasons why evaluation is important to look for in teacher discussions in an online context where the main communication channel is text. Indeed, in systemic functional linguistics (SFL), evaluation is the function accomplished by the elements of the *appraisal system of language* (Martin & White, 2005). Appraisal, in turn, is one of the systems of resources with which language realizes what Halliday calls the *interpersonal* metafunction of language (Halliday & Matthiessen, 2004), or how language enables speakers and writers to relate to their audience, particularly engaging the audience (Martin, 2000). Appraisal theory would predict that more evaluation tokens in a text mean more attempts to engage readers or listeners of the text. It could be argued that more engagement of writers and readers contributes to a sense of social presence in the forum; that is, the existence of tokens of evaluation in a posting could indicate that participants experience the forum participation as a real conversation with others.

¹ A reviewer pointed out that it is likely that the recommendations in the literature that caution against evaluation are directed to evaluation of teachers (as individuals) rather than to evaluation of teaching (as a professional practice). We acknowledge this possibility. This reviewer also noted that in coaching, it is particularly important for coaches not to be evaluators (Tschannen-Moran & Tschannen-Moran, 2011).

Thus, while the literature documents the importance of reflection and alternativity in forum conversations and other teacher development encounters, the value of evaluation seems controversial. A study of the relationships between evaluation, reflection, and alternativity seems productive to help understand participants' sense of social presence in teacher forums.

This paper investigates whether making evaluative comments correlates with making reflective comments and with proposing alternative teaching actions. We examine this question in the context of *online* forums hosted in the LessonSketch platform (see Herbst, Aaron, & Chieu, 2013), which we describe in more detail in the next section. These online forums used animated representations of classroom episodes, where teacher and students had been represented with nondescript cartoon characters. There has been little research in the literature on the use of animations in teacher development (Herbst et al., 2011; Moore-Russo & Viglietti, 2011; Moreno & Ortegado-Layne, 2008; Tettegah et al., 2008). An important advantage of the use of nondescript cartoon-based representations of teaching is that it makes it easier for the audience to focus on practices (as opposed to focusing on the individualities of people and settings) and more comfortable for the audience to appraise the cartoon characters' actions. An earlier study (Kosko & Herbst, 2012; Herbst & Kosko, in press) indicated that while in general participants' levels of modality usage, which are included in the appraisal system in SFL, were similar when annotating either videos or animations, their level of normativity usage (e.g., "the teacher should ...") was significantly higher when they watched and discussed animations than when they watched and discussed videos. If evaluative comments were connected to increasing reflection and alternativity, animations might offer a useful alternative to video, about which Seago (2004, p. 275) has noted that when teachers discuss video records "politeness and agreement is the norm" and that teachers tend to handle differences with comments such as

“everybody needs to teach according to his style.” But, as noted above, we consider a key feature of learning from practice that participants be able to propose alternative actions and to reflect on instructional practice, and we conjecture that both of them correlate with some degree of appraisal.

LessonSketch: A Web-based Interactive Rich-Media Environment for Teacher Learning

LessonSketch is a web-based, interactive rich-media environment that supports collaborative learning for teachers (Chieu & Herbst, 2012; Herbst, Aaron, & Chieu, 2013; Herbst, Chazan, Chen, Chieu, & Weiss, 2011). Its design has been grounded in activity theory (Engeström, 1999; Kaptelinin & Nardi, 2006; Leontiev, 1978; Vygotsky, 1978) and practice-based perspectives on teacher education (Ball & Cohen, 1999; Grossman et al., 2009; Lampert, 2010). A key characteristic that makes *LessonSketch* stand out among video-based learning environments for teachers is the use of cartoon-based representations of teaching: animations and storyboards in which cartoon characters represent scenarios of classroom interaction (Herbst & Chazan, 2006; Herbst et al., 2011; Herbst & Miyakawa, 2008). The use of stylized cartoon characters can help design and create representations of teaching very flexibly. These representations can profit from some of the advantages of written narrative cases, such as the possibility to represent with icons the individualities and settings involved, and thus control how much of those are relayed to, versus evoked from, the audience. Animations of cartoon characters can also profit from some of the advantages of video cases, such as the possibility to communicate multimodally (e.g., using gesture, facial expression, and body movement and position in addition to language) and to involve the audience in a temporality (i.e., a sense of how time flows) and timeliness (i.e., a sense that actions happen at the moment when they are needed) commensurate with that of real

action (Herbst et al., 2011) and without video's "multiple channels of distractions" (Goldsmith & Seago, 2011, p. 184).

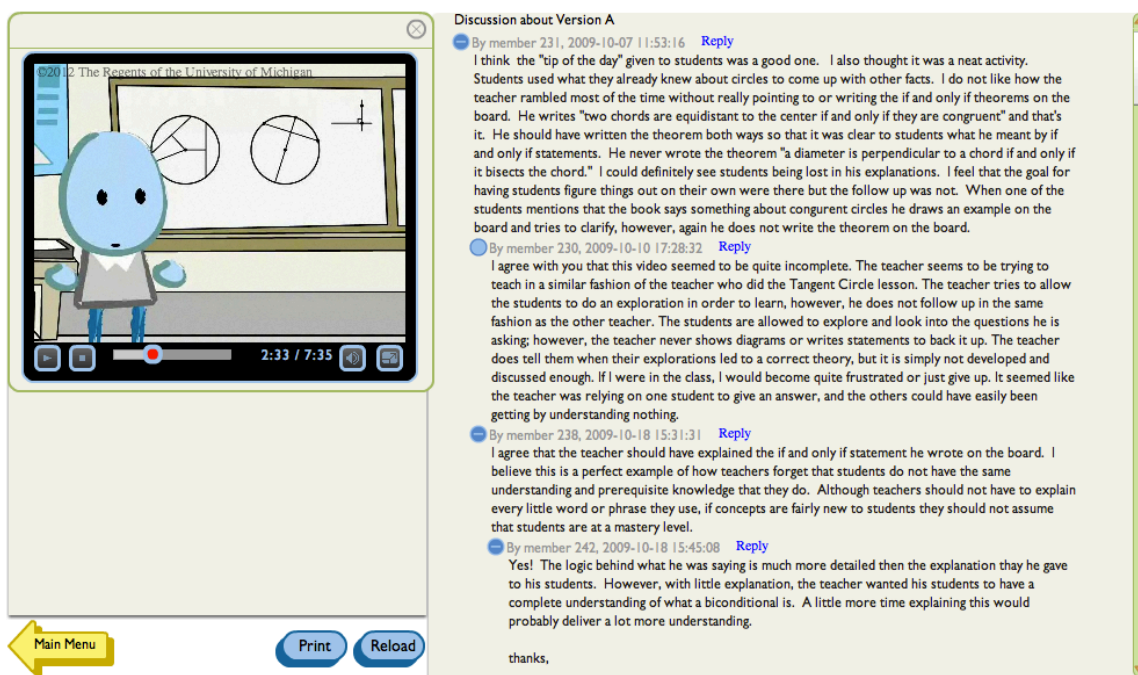


Figure 1. Part of LessonSketch's communication tools: embedding an animation into the discussion space (users' name is replaced with a system-generated id to protect their identity).

LessonSketch's users can use advanced communication tools to engage in collective and collaborative reflection. Unlike traditional and text-based communication tools (Barab, et al., 2004; Fishman & Davis, 2006), LessonSketch can embed representations of teaching (e.g., animations, storyboards, videos), as shared artifacts, into the virtual discussion space to enhance online professional conversations (see Figure 1 for an example), because when referring to shared artifacts and focusing on learning contents, participants are more likely to produce meaningful, in-depth discussions (Chieu, et al., 2011; Neale, Carroll, & Rosson, 2004; Wise, Padmanabhan, & Duffy, 2009). We contend that LessonSketch's cartoon-based artifacts and tools play a crucial role, as "mediators of cognition," to help teacher users externalize their thoughts and ideas about instructional practice, and thus develop shared goals and

understandings (Engeström, 1999). This paper partly investigates the nature of online discussions through *LessonSketch*'s advanced communication tools.

Research Design

Research Questions

We conjecture that when teacher users evaluate classroom events (i.e., when they explicitly make evaluative comments on classroom events within a public conversation space), they are more likely to propose alternative actions for the animated teacher or students and to interpret or reflect on their professional practice than when they do not evaluate classroom events. An important goal for research has been to verify the presence of those qualities in collaborative, professional learning activities. In this study, we thus consider *reflection* and *alternativity* as two key desirable features of teacher discourse. The main research question of this study is to investigate the correlations between *evaluation*, as an indicator, and those two discursive features. More specifically, we focus on the following questions: Are there any associations between the observation that participants evaluate events of embedded animations and (a) the observation that they anticipate alternative actions by teacher or students, and (b) the observation that they interpret or reflect on what they notice? What is the nature of those associations, if any? Are there any significant effects of the way individual participants post in forums (i.e., the frequency of their forum posting) on those correlations?

Settings, Participants, and Procedure

In the fall of 2009, a mathematics teacher educator at a university in the Eastern United States asked us to create eight online sessions for a class on geometry instruction. Each online session

was a structured exploration and discussion of one or several versions of animated classroom stories (each story has a main branch and sometimes a number of variations, for example, with alternative endings). More specifically, it consisted of the following consecutive activities: (a) an individual exploration in which participants were asked to view and comment on one or more versions of an animated story, and (b) a forum discussion of those animations. We created a discussion thread for each animation version. The user interface of those threads was similar to the one presented in Figure 1; the main feature included an animation that was directly embedded on the left hand side of the discussion space, which was organized in a tree-based format. Figure 1 shows the discussion of Version A (a main branch) of the “Chords and Distances” story in which the teacher asks students to work in groups to form conjectures about circles, chords, and their distance to the center of the circle. Table 1 shows the use of animated stories over the one-semester class. One session was used in each of the eight weeks; all participants took part in the same session at any given time (i.e., there was only one group of participants in this study); participants usually had almost the whole week to post messages after a face-to-face class meeting every Monday. Twenty-one participants (11 teacher candidates and 10 novice teachers, 16 females and 5 males) enrolled in the course. None of the teacher candidates had full-time classroom teaching experience, though some had temporary teaching experience. The novice teachers had a full-time teaching job, with no more than three years of teaching experience each. The forums were not moderated; the teacher educator read participants’ postings but did not make any comments or gave feedback on those postings; she did not encourage or discourage evaluation by teachers. The participants were informed that the teacher educator would grade the thoughtfulness and insight of their forum postings at the end of

the course. But the teacher educator was not involved in the present study, analyzing data or interpreting or reporting results.

Table 1. Use of animation versions in eight weekly online sessions.

Week	Date	Story Title	Version
1	2009-09-14	Parallel Lines	A
			B
			C
2	2009-09-21	The Isosceles Triangle	C
			D
3	2009-10-05	Chords and Distances	A
		The Tangle Circle	A
4	2009-10-12	Intersection of Medians	A
5	2009-10-19	The Parallelogram	A
		The Square	A
6	2009-10-26	A Proof about Rectangles	D
7	2009-11-09	The Kite	A
8	2009-11-16	The Kite	E
		The Midpoint Quadrilateral	B

In the design of animated classroom scenarios, for each instructional story we created a number of critical events to prompt participants' conversations about teaching practice; by critical events, we mean moments in which instructional norms are breached (Herbst & Chazan, 2003; Herbst, et al., 2011). For example, a breach of a norm of how proof tasks in an American

high school geometry lesson are assigned could be instantiated if when giving students a problem where they are expected to produce a proof, the teacher did not provide students with clear statements of the givens and conclusion to prove; the norm is that the teacher will do so (Herbst, Aaron, Dimmel, & Erickson, 2013).

Data Sources and Data Analysis

To respond to the questions mentioned above, we collected all forum logs. As a rule, we took the posting as the unit of analysis, making the assumption that each posting contained a single contribution to the discussion. Each posting was coded for the presence or absence of characteristics of interest. The only exception was the case in which a posting included more than one paragraph and there were explicit markers of contrast in the form of internal conjunctions, such as “on the other hand,” to connect the paragraphs. Those markers suggest that the posting might include more than one contribution to the discussion. Each time such a marker existed, we considered a new unit of analysis. This is reasonable because some members may prefer to use that kind of marker to connect ideas in one posting instead of separating ideas into multiple postings.

We used elements of systemic functional linguistics (SFL; Halliday & Matthiessen, 2004; Martin & Rose, 2007) to code text-based conversations in forums (we describe this coding in the next sub-section). SFL provides the basis for an operational framework with which we can identify, for example, where participants made evaluations of teaching or where they reflected on the work of teaching. SFL has been increasingly used in education research (Martin, 2001; Schleppegrell, 2012a, 2012b).

Because, as a rule, each individual made several postings, we did not consider postings as independent of each other. Thus, we used Hierarchical Generalized Linear Modeling (HGLM, described below), a particular form of multilevel modeling, to handle the structure that postings were nested in individual participants. Multilevel models are powerful ways to deal with the nested structure of data (e.g., students nested in classrooms, classrooms nested in schools, schools nested in districts). Recently, a number of studies have used multilevel models to analyze data in online discussions (see Cress, 2008, for a more extensive review). Note that while individuals are typically nested in groups for multilevel models, and thus a three-level model (postings are nested in individuals and individuals nested in groups) would be more appropriate to analyze conversation data, since we had only one group we did not consider the individual-group nesting structure in our analysis. In addition, postings are also nested in discussion threads that participants may create in forums and threads are nested in forums. Because the participants in this study used only a small number of threads and forums that the teacher educator initiated in advance, however, we did not include crossed random effects of individuals and threads in our models and we did not consider forums as a new level in the nested structure of data. More explicitly, our sample included 21 forum participants making 723 posts over an eight-week period. Posts were made in response to parent posts from the instructor, and while it would have been ideal to account both for posts being made from individuals and in response to particular posts, we were limited by our sample size. Our decision to account for postings as nested within individuals allowed for an examination of important features of the online discussion. However, we acknowledge the limitation that our analysis does not statistically examine the effects of postings as also nested within parent posts².

² It is possible to statistically account for the dual nesting effects of parent post and individual. However, this

While our statistical analysis does not fully account for the turn-based nature of discourse, our consideration of postings as nested in individual participants makes the assumption that while one's posting may be influenced in some fashion by those in a thread, it is the individual who is ultimately responsible for what to put in a posting. While there are limitations in assuming that postings depend only on individuals who make the postings, our assumption is not unlike what is often assumed about classroom discourse where individual students are assumed responsible for their actions and words, while students in fact interact with one another and thus such interactions might also influence the manner in which individuals perform in those classrooms. Just as it is acceptable to assume a reasonable level of independence for the sake of parsimony in such cases (students nested within classrooms), we apply the same logic to the case of online forums (postings nested within individuals). This is not to say that examination of the turn-based, interactive nature of discourse is irrelevant; in fact, we consider it important to explore, but given the nature and size of our present sample, we take postings as nested within individuals as a reasonable approximation of the phenomenon. Although we were not able to include both nesting structures (posts nested within individuals and posts nested within parent posts) in a single HGLM model, we performed another similar analysis in a preliminary study (Chieu & Herbst, in press) to investigate correlations between the presence of *evaluation* markers in a parent post and the presence of *reflection* or *alternativity* markers in a follow-up post.

Systemic Functional Linguistics and Coding Scheme

We incorporated SFL to code for lexical and grammatical elements in forum postings (Halliday & Matthiessen, 2004; Martin & Rose, 2007). We tracked uses of reference and substitution to

requires a process referred to as crossed random effects for which we did not have a sufficient sample size (accounting for the number of parent posts, posts, and individuals posting).

detect cohesion in online discussions when needed (Schiffrin, Tannen, & Hamilton, 2003). SFL, which considers language as a social semiotic system, looks at the language choices people make to construe meaning: It describes in particular the grammatical and lexical choices available to construct discourse and the meanings that are constructed through those choices. One can better understand how people construct meanings by contrasting the options they choose against the other options they could have chosen.

According to Halliday (see Halliday & Matthiessen, 2004), speakers convey meaning by simultaneously drawing upon the resources that language has available to fulfill three fundamental meta-functions of language: First, the *ideational* meta-function has to do with the language resources for construing experiences or ideas. Second, the *interpersonal* meta-function is concerned with the resources that language provides for creating and maintaining social relations between speaker or writer and listener or reader. Third, the *textual* meta-function is concerned with the resources that language has available to *organize* its products into texts of particular genres. Each text is composed of linguistic choices that perform those three metafunctions simultaneously.

Our coding scheme sought to linguistically track the properties of conversations about the animations identified by Herbst and Chazan (2006) through the observation of face-to-face study groups. Our coding scheme also sought to linguistically track part of the codes (e.g., evaluative stance) used by van Es and Sherin (2008) and part of the codes (e.g., alternativity) we had proposed in an earlier study of *LessonSketch* (Chieu, et al., 2011). This coding system attends to the three variables mentioned previously: *evaluation*, *reflection*, and *alternativity*. Improving upon earlier usages of those codes, we operationalized those codes through attention to the linguistic choices participants made in the forums. Two coders coded postings independently.

Any doubtful cases, where the relevance to teaching practice was not obvious, were not coded. Aided by an SFL-based operationalization of the three codes, the two coders first coded 20 units independently. Then, they reconciled the two analyses and revisited the coding process. They repeated the same procedure for another set of 20 units. When the two coders believed that the coding system was reliable, they independently coded one forum log (about 50 units). We used the Cohen's Kappa statistic to determine inter-rater reliability. The kappas of the first coding round indicated a moderate reliability, so the two coders reconciled all differing codes, continued to improve the coding scheme and their coding skill, and independently coded another forum log (about 50 units). Cohen's kappa statistics in the second coding round for *evaluation* was .66, that for *alternativity* was .77, and that for *reflection* was .69 ($p < .001$). Kappas ranging from .01 to .20 are considered to have slight agreement, .21 to .40 are fair, .41 to .60 are moderate, .61 to .80 are substantial, and .81 to 1.00 are almost perfect (Sim & Wright, 2005). Therefore, the scores obtained for our coding suggest good inter-rater reliability (Capozzoli, McSweeney, & Sinha, 1999; Sim & Wright, 2005). Finally, the two coders reconciled all differing codes and continued to code the remaining units, half for each coder. We describe below how SFL helped us operationalize the three codes.

For the *evaluation* code, we used Martin and White's (2005) appraisal theory, which develops the systemic functional approach to describe the appraisal system as composed of the subsystems of *affect*, *judgment*, and *appreciation*.³ That approach helped us identify where participants made evaluations of teaching. According to Martin and White (2005), appraisals

³ These subsystems are listed to flesh out what appraisal means. Appraisal can be expressed as a feeling of the speaker (i.e., affect; e.g., "that made me sick") or institutionalized, either as a judgment of the people involved (e.g., "she is incompetent") or as an appreciation of the goods or services involved (e.g., "that was a terrible explanation"). These distinctions helped us recognize appraisal, though the coded data was not disaggregated among those categories.

“...reveal the speaker’s/writer’s feelings and values...” (p. 2) and can be realized not only lexically (through word choice) but also grammatically (e.g., through the use of modals such as *should* in “she should not have...”). We counted the number of the participants’ evaluations of teacher, students, objects, and actions in the animated classroom. A posting was coded 1 for *evaluation* if it contained at least one marker of such appraisal, and 0 otherwise. Table 2 shows examples of appraisal markers (adapted from Martin & White, 2005, Chapter 2).

Table 2. Examples of markers for codes.

Code	Type of Marker	Examples
<i>Evaluation</i>	affect markers	indications of how participants felt such as <i>cheerful, like, confident, comfortable with, curious, satisfied</i> (positive) or <i>sad, dislike, hate, anxious, surprised, bored, angry</i> (negative)
	judgment markers	indications of how participants assessed people in the animation such as <i>clever, tireless, reliable, direct, good</i> (positive) or <i>odd, weak, slow, unreliable, bad, unfair, mean</i> (negative)
	appreciation markers	indications of how they assessed actions in the animation such as <i>engaging, lovely, exciting, beautiful, appealing, balanced, unified, simple, elegant, rich, challenging, innovative, original, unique</i> (positive) or <i>boring, ugly, unbalanced, simplistic, insignificant</i> (negative)
<i>Reflection</i>	causal-conditional conjunctions	enhancement that modifies clauses through variations of logical connections, for instance, <i>because, as, since, so that, if (then), unless, without</i>
	manner or means or comparisons	enhancement that qualifies meaning through comparison or the means in which the process of one clause is enacted, for example, <i>and thus, and so, by (means of), instead of, which means that, to (in order to), however</i>
<i>Alternativity</i>	use of modals (could, should)	They <i>could</i> work in groups.
	subjunctive mood	<i>If</i> the teacher <i>provided</i> the givens and prove.
	potential mood	They <i>would</i> like another problem.
	negative use of indicative mood	The teacher <i>did not</i> provide students with the givens and prove.

For the *reflection* code, we used Halliday and Matthiessen's (2004) notion of clause enhancement, with specific attention to *manner* and *causal-conditional* enhancements. According to Halliday and Matthiessen (2004), "in enhancement, one clause [...] enhances the meaning of another by qualifying it in one of a number of possible ways" (p. 410). *Manner* enhancement qualifies meaning through comparison or the means in which the process of one clause is enacted. In the example "[Students] could check their own logic and reasons by putting [the proof] into a two column form", the clause "[Students] could check their own logic and reasons" is enhanced with the information that this is done through using a two-column proof format. *Causal-conditional* enhancement modifies clauses through variations of logical connections (e.g., if P then Q; because of Q, so P, etc.). The example "So I don't think we should teach to the test because it isn't necessary" presents this form of enhancement where rationales are provided in varying orders and formats with indicators such as "so" and "because." Both forms of enhancement (manner and causal-conditional) were taken as evidence of reflection due to their demonstration of logical reasoning. Providing rationales or means for which actions take place (in the form of grammatical processes in a clause) is evidence of thinking about thinking, and therefore characteristic of reflection. A posting was coded 1 for *reflection* if it contained at least one marker of reflection on teaching practice, and 0 otherwise. Table 2 provides examples of reflection markers (adapted from Halliday & Matthiessen, 2004, Section 7.4.3).

For the *alternativity* code, we counted the presence of teaching actions that should, could, or would have been taken in the animation. A proxy for alternativity could be the participant's use of modals (could, should, would) in reporting the proposed action, but not necessarily. We used grammatical mood to determine whether or not the participants were talking about events that had happened or had not happened. In any case, the coder should be able to point out what

the original action and the alternative action were. A posting was coded 1 for *alternativity* if there was at least one marker of it, and 0 otherwise. Table 2 shows examples of alternativity markers.

We give an example of the final codes for a couple of typical postings in Table 3. Note that in addition to examining markers in a posting, for each forum the coder watched the embedded animation to understand the content of the classroom story the participants were talking about. The coder also looked at the parent postings (i.e., the posting that the posting being analyzed replied to) to make better sense of the context of the posting.

Table 3. Assignment of codes of two typical postings in Week 4 (User 230 replied to User 231).

We underlined pieces of the text that indicate the presence of the code in brackets.

User 231 (Post ID = 684): I like [EVALUATION] how the teacher talks about the triangle's center of gravity and the example that he uses (cardboard and balancing it on a pencil). This is a nice [EVALUATION] lead into the three smaller triangles being equal in area. I also think it is good [EVALUATION] that he discusses that only two of the medians need to be drawn because [REFLECTION] we already know that the third will pass through O. However [REFLECTION], it would have been better if he had made [ALTERNATIVITY] the connection between this and the statement that he had written on the board earlier about the medians of a triangle meeting at a unique point. He never discusses the word unique [EVALUATION, ALTERNATIVITY]. I really liked [EVALUATION] that he has the students justify to themselves that the theorem is true by measuring. I think this solidifies the concept for them rather than just having it in words [EVALUATION]. I think students would be more likely to believe that it is true in doing this. However [REFLECTION], we see that there is a typical class ending. The teacher is trying to apply what they have learned and the students do not seem to make the connection between it and what they have just done. They say that the triangles will be congruent (not equal in area) and then when one finally says that they would be equal in area another yells "No way! One of them is fat and the other is skinny." They just did an example where the triangle [sic] different shape but still had the same area. So [REFLECTION] my question is...do they really understand what they just proved?

User 230 (Post ID = 697): I think a way around this would be to have the students think, pair, share their ideas and answers [ALTERNATIVITY]. If they were working on it independently, they may have understood the connections better because [REFLECTION] they were working through it themselves. As a class, as they are doing it here, some students are working harder at the problem than others. Also, the teacher makes some statements without explaining [EVALUATION], which can be corrected by the sharing portion where the students explain what they are thinking and how it relates to past ideas [ALTERNATIVITY, REFLECTION].

Multilevel Models

We used Hierarchical Generalized Linear Model (HGLM) to examine the probability for forum postings to include alternativity and reflection. HGLM is a non-parametric analysis of multi-level data and, therefore, it calculates logits to estimate the likelihood of certain outcomes with given conditions. These estimations are generally limited to the samples examined. Within this study, our findings are limited to the participants engaged in the online forum and our estimates concern the correlation between the likelihood of certain elements in a post with the likelihood of other elements being included. HGLM allows for the examination of nested data (postings nested within the individual) by creating regression equations for each level of analysis and using the slopes of lower-level regressions as the outcome measures for higher-level regressions. Essentially, HGLM allowed us to create an identical logistic regression equation for each individual participant and for all postings of that individual as the unit of analysis for such regression equations (logistic regression was used to model dichotomous outcome variables). Using the HLM 6 program (Raudenbush, Bryk, & Congdon, 2004), we calculated statistical effects in the form of logits for each variable serving as an indicator, while adjusting for the variance within each individual who posted in the forum. These logits were used as outcome variables (or those discursive features indicated by *evaluation*) for regression equations examining individual characteristics. We describe this model in detail in the following paragraphs (see Chapter 10 of Raudenbush & Bryk, 2002, for a complete description of the method).

We looked at the correlations between making an *evaluation* of the teaching in the animation within a posting and the probability that the posting contained *reflection* (viz.,

alternativity). The proposed final model is outlined, for the case of *reflection* (viz., *alternativity*) as the outcome⁴ and *evaluation* as an indicator of the probability that *reflection* is present. However, variations using other variables are identical, and we limit our description of the models to the example below for simplicity:

Level-1 Equation:

$$reflection_{ij} = \pi_{ij} \quad \pi \sim \text{Binomial}(\eta_{ij}, \mu)$$

$$\pi_{ij} = \text{logistic}(\beta_{0j} + \beta_{1j} * evaluation_{ij} + \mu_{ij})$$

Level-2 Equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * N_Posts_j + \gamma_{02} * Status_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * N_Posts_j + \gamma_{12} * Status_j + u_{1j}$$

The outcome measure of the level-1 equation, π_{ij} , represents the transformed predicted value for whether or not *reflection* is present in forum posting i , made by individual j . At level-1, error is measured by μ_{ij} for posting i nested within individual j . At level-2, error is measured by u_{0j} and u_{1j} for both parameters for individual j . The intercept, β_{0j} , and the effect of the presence of *evaluation*, β_{1j} , for participant j , are expressed in logit units, which are calculated through Maximum Likelihood Estimation (MLE). β_{0j} represents the average logits for individual j across all postings made by that individual. Similarly, β_{1j} represents the average effect-size for

⁴ Note that while the statistical model is described using words like indicator and outcome (or independent and dependent variable), this does not mean an assumption is being made of any causality between the phenomena described by those variables.

individual j across all postings. Once these beta values are calculated, the probability of *reflection* in a forum post i for individual j can be calculated with the following equation:

$$P(\text{OutcomeReflection}) = \frac{e^{\beta_{0j} + \beta_{1j} * \text{evaluation}}}{1 + e^{\beta_{0j} + \beta_{1j} * \text{evaluation}}}$$

The coefficients at level-1 are estimated as outcomes at level-2, which represents data of the individual forum participant. By estimating the effects of level-1 as outcomes at level-2, we were able to account for some of the variance within individual participants, for example, their posting habits. We examined the effects, γ_{01} and γ_{11} , of how frequently or actively participants made forum postings, and included the variable N_Posts (i.e., the total number of postings an individual made in all forums), grand mean centered at level-2. Thus, a unit increase or decrease in N_Posts in the level-2 equation represents an increase or decrease with respect to the overall mean posting frequency of all participants. We also included *Status* as another level-2 variable, to account for whether a participant was a teacher candidate or prospective teacher (no teaching experience) or a novice teacher (no more than 3 years of classroom teaching experience), which were the two kinds of participants in the forums examined. *Status* was dummy coded (0 = prospective teacher; 1 = novice teacher) and its effects represented by γ_{02} and γ_{12} in the model. This allows different effect sizes to be used for different individuals such that two participants with different posting frequencies and different status would have different effect sizes for their forum postings at level-1. The construction of the models and results of the HGLM analysis are presented and discussed for each specific model in the following section.

Results and Discussion

Overall, 21 participants contributed 723 postings over 8 online forum sessions. The mean number of posts per participant was $M = 34.4$, with $Min = 21$ and $Max = 46$, indicating a power coefficient of approximately .88, which was updated using Optimal Design 1.83 (Liu, Spybrook, Martinez, & Raudenbush, 2007). While the estimated power is sufficient for our analysis, it is important to consider the results that follow in context. Specifically, the sample includes 723 postings, but these come from only 21 participants. Thus, we do not make any claims of generalizability of our findings, but only regarding trends in our particular sample.

Participants often made evaluations of the animations (Figure 2) and reflected on what they noticed (Figure 3). They proposed a large number of alternative actions for the instructional practice represented in the animations (Figure 4). In other words, forum conversations in each of the eight online sessions were highly interactive and all members actively contributed to those valuable discussions (see also Chieu & Herbst, 2011). The offline class activities might have helped stimulate that level of interaction. We believe, however, that the embedded animations may have also contributed to that phenomenon, because teachers' discussion was highly interactive and meaningful even on the first week, and that phenomenon was also recognized in single-session studies that we conducted earlier (Chieu, et al., 2011).

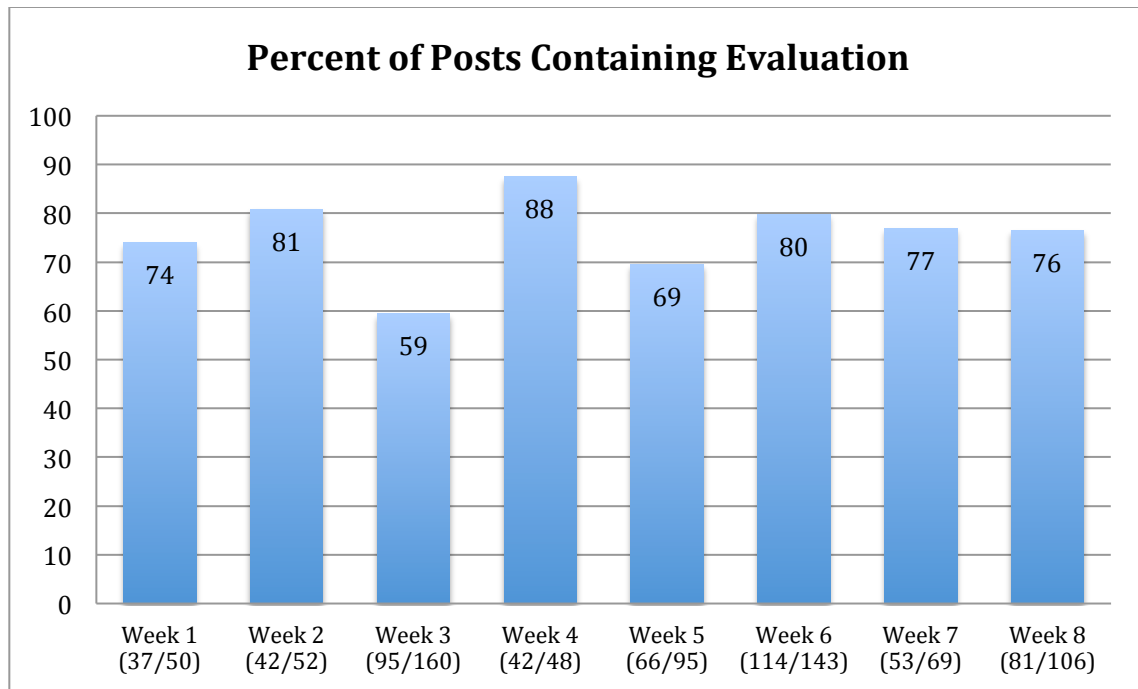


Figure 2. Percentage of posts that include evaluative comments over 8 weeks (for each pair of numbers in the parentheses below the week label, the first number indicates the number of posts containing evaluation and the second number indicates the total number of posts of the week).

The examination of the coded data allowed us to notice that the presence of *evaluation* often concurred with the presence of *alternativity* and *reflection* in a posting (see examples in Table 3). Next, we investigate more about those correlations using HGLM, and we show that evaluating features of classroom interactions significantly correlated with the participants' reflection on instructional practice and their proposal of alternative actions in teaching.

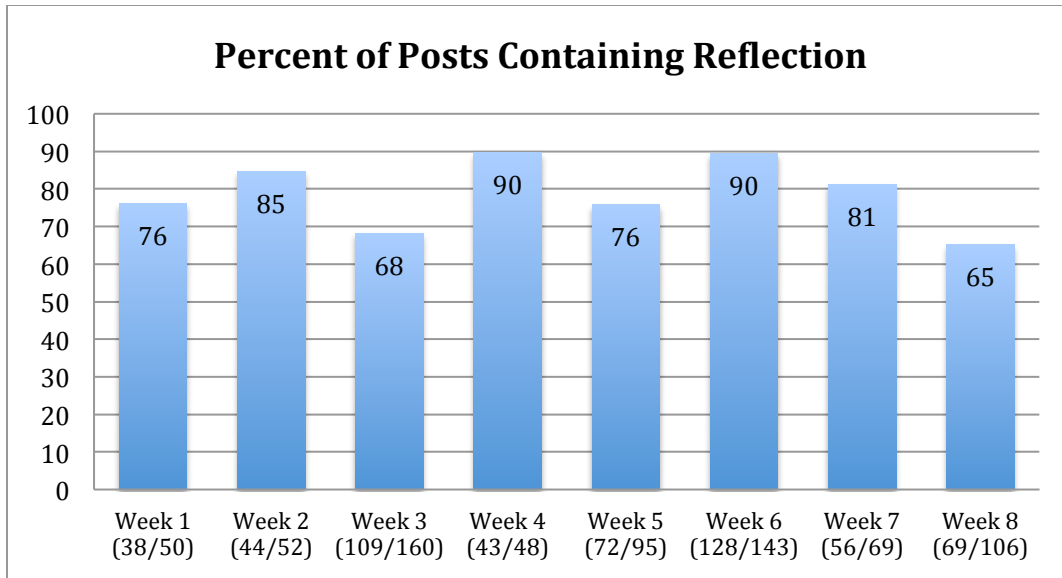


Figure 3. Percentage of posts that include reflective comments over 8 weeks (for each pair of numbers in the parentheses below the week label, the first number indicates the number of posts containing reflection and the second number indicates the total number of posts of the week).

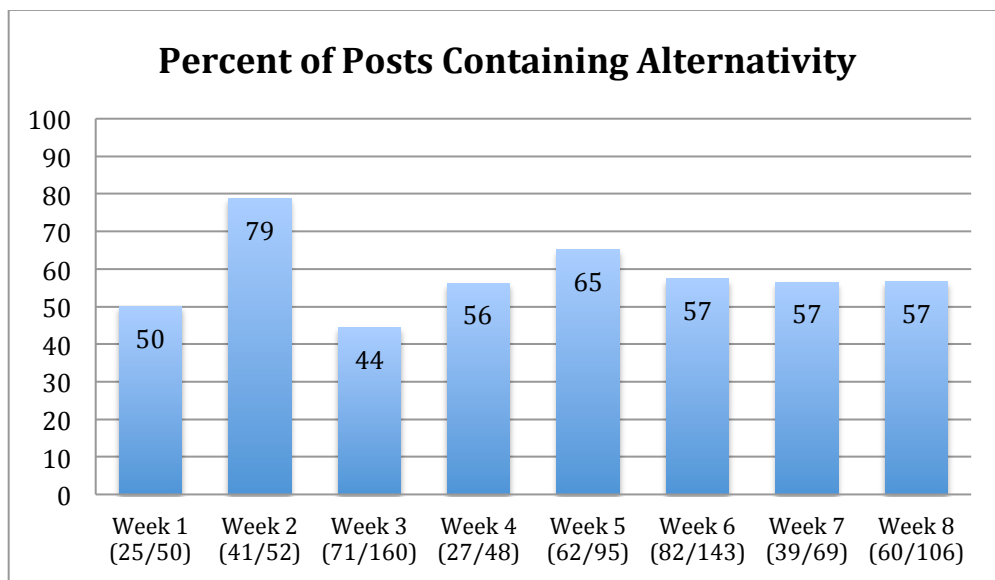


Figure 4. Percentage of posts that include proposals of alternative teaching actions over 8 weeks (for each pair of numbers in the parentheses below the week label, the first number indicates the number of posts containing alternativity and the second number indicates the total number of posts of the week).

Analysis of Reflection

Table 4 summarizes the frequencies of *evaluation* and *reflection*. Table 4 shows a dominance of the frequency of *evaluation* = 1 and *reflection* = 1. Overall, about 77% comments included reflection. If counting only comments including evaluation, however, that percentage increased to about 83%. The following HGLM analysis provides a better picture of the correlation between those two variables, by accounting for the nested structure of the data (i.e., posts made by individuals).

Table 4. Frequencies of Evaluation and Reflection.

	Evaluation		Total
	0	1	
0	73 (38%) (45%)	120 (62%) (21%)	193 (100%) (27%)
1	91 (17%) (55%)	439 (83%) (79%)	530 (100%) (73%)
Total	164 (23%) (100%)	559 (77%) (100%)	723 (100%) (100%)

Customary in hierarchical linear modeling is the construction of models from their more basic elements to the final proposed model (Hox, 2002; Raudenbush & Bryk, 2002), such as that presented in the previous section. A first step in this process is the construction and running of an unconditional or empty model. The unconditional model contains only the outcome measure (in this case, *reflection*). Therefore, the intercept, γ_{00} , represents the average logit for all participants in the sample and the resulting outcome for the model provides the average probability for a posting made by any participant to contain reflection:

Level-1 Equation:

$$reflection_{ij} = \pi_{ij} \quad \pi \sim \text{Binomial}(\eta_{ij}, \mu)$$

$$\pi_{ij} = \text{logistic}(\beta_{0j} + \mu_{ij})$$

Level-2 Equation:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Results indicated that the intercept was statistically different from zero ($\gamma_{00} = 1.33$, $p < .001$), suggesting that individuals were more likely to make forum postings with reflection than not. The intra-class correlation coefficient was statistically different from zero (ICC = .10, $p < .01$). This means that around 10 percent of the variance in *reflection* estimation was due to differences across individuals ($p < .01$), with the remaining 90 percent attributable to post differences. So, we examined the full multilevel model described previously to see if *Status* and/or *N_Posts* could play a critical role in those differences across individuals.

In constructing the model including *evaluation*, we first constructed a baseline model (model 1), which included only the level-1 variable *evaluation*. Results showed a reduced size for the intercept from the unconditional model ($\gamma_{00} = .54$, $p = .011$), and a statistically significant effect for *evaluation* ($\gamma_{10} = 1.16$, $p < .001$). These results (Figure 5) suggested that a forum posting that did not contain an evaluation of the teaching in the animation had a 63.2% chance of including reflection. However, if the posting did contain evaluation, then the probability of including reflection improved to 84.6% (effect size or odds ratio = 3.2, $p < .001$). In other words,

the odds of a comment including reflection if it did contain evaluation were 3.2 times higher than the odds of a comment including reflection if it did not contain evaluation.

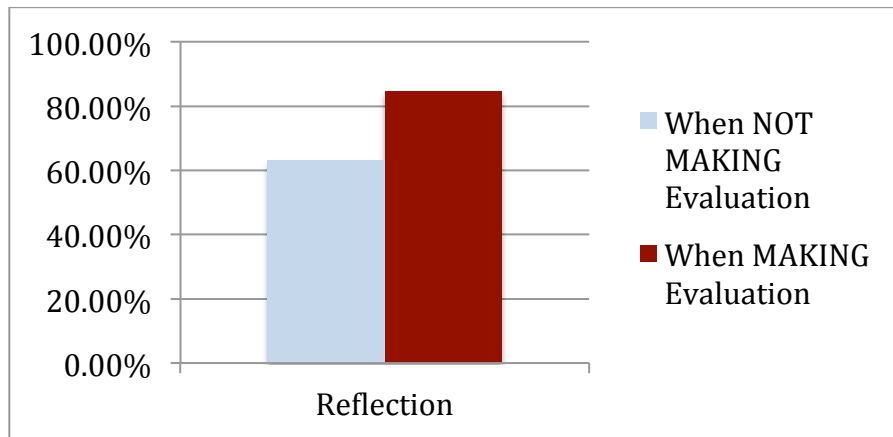


Figure 5. Probabilities of making reflective comments.

Model 2 included the variables N_Posts and $Status$ as variables at level-2. Results of this model indicated that an individual's status as either a future teacher or a novice teacher had no statistically significant interactions with the level-1 intercept ($\gamma_{02} = .23, p = .59$) or with the effect of the level-1 variable ($\gamma_{12} = -.12, p = .84$). Similarly, an individual's total number of posts in all forums had no statistically significant interactions with the level-1 intercept ($\gamma_{02} = -.03, p = .47$) or with the effect of the level-1 variable ($\gamma_{12} = .04, p = .4$). So, although about 10% of the variance in *reflection* estimation was due to differences across individuals, the participants' status and their total number of posts did not contribute to those differences significantly. Given that our sample included only 21 participants, it may be that a larger sample would find differing results. In such context it would be useful to have available other individual measures that might account for the variance among individuals that we could not account for here by attending to forum behavior: For example, participants' level of mathematical knowledge for teaching might help explain some of this variance. In summary, we found that model 1 described above was the

best-fit model for the current set of data. This model suggested that *evaluation* had a strong statistical effect on the probability of a *reflection* in a posting.

Analysis of Alternativity

Table 5 summarizes the frequencies of *evaluation* and *alternativity*. Table 5 shows a dominance of the frequency of *evaluation* = 1 and *alternativity* = 1. Overall, about 56% comments included alternativity. If counting only comments including evaluation, however, that percentage increased to about 65%. We applied the same analysis process described above to investigate the correlation between those two variables.

Table 5. Frequencies of Evaluation and Alternativity.

Evaluation	Alternativity		Total
	0	1	
0	132 (68%) (42%)	61 (32%) (15%)	193 (100%) (27%)
1	184 (35%) (58%)	346 (65%) (85%)	530 (100%) (73%)
Total	316 (44%) (100%)	407 (56%) (100%)	723 (100%) (100%)

The unconditional or empty model indicated that individuals were more likely to propose alternative teaching actions in forum postings than not: the intercept was statistically different from zero ($\gamma_{00} = .26, p = .002$). The intra-class correlation coefficient, however, was not statistically different from zero ($ICC = .008, p = .25$), meaning that the differences across individuals did not contribute to the variance in *alternativity* estimation significantly. Thus, we

used a logistic regression model for this analysis (Cress, 2008). The results (Figure 6) of this model suggested that a forum posting that did not contain an evaluation of teaching in the animation had a 31.6% chance of including alternativity. Yet, if the posting did contain evaluation, then the probability of including alternativity increased to 65.3% (effect size or odds ratio = 4.1, $p < .001$). In other words, the odds of a comment including alternativity if it did contain evaluation were 4.1 times higher than the odds of a comment including alternativity if it did not contain evaluation. Similarly to the analysis of reflection, we found that *evaluation* had a strong statistical effect on *alternativity*.

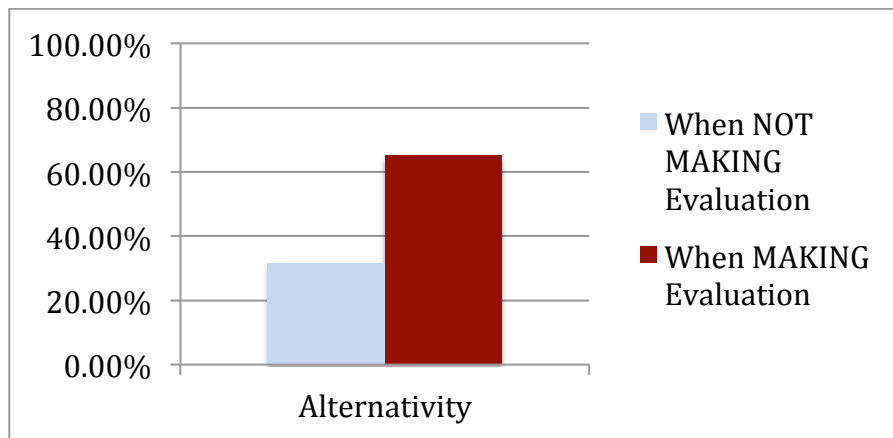


Figure 6. Probabilities of proposing alternative teaching actions.

Discussion of all Correlations

The previous analyses suggest strong correlations between evaluation and reflection and between evaluation and alternativity in online discussions by candidate and novice teachers. From our observation while coding forum posts, we found that the quality of participants' reflection and alternativity, regarding teaching practice, throughout all discussions was relatively high. Yet posts that contained evaluation were more likely to have those desirable characteristics than posts that did not contain evaluation. This result seems to suggest that discouraging participants

from making evaluative comments might not be conducive to improving the quality of discussions, especially in discussions where the reference object is not a video from one of the participants' own teaching. Of course, it is important to replicate this result in further studies before making recommendations that might be consequential.

The importance of looking for those correlations was justified on the role that evaluation plays in supporting the construction of interpersonal relationships through language, as suggested by systemic functional linguistics. Along those lines, the conjecture was that a forum where participants were not discouraged from evaluating could enable (or not disable) resources that might support constructing an online asynchronous discussion that felt more like a conversation among people. The results show that when speakers (or rather, forum contributors) engage those resources, they are also more likely to contribute content that could be considered valuable (on account of including reflection and proposing of alternatives). A question for further research, however, is whether the use of appraisal resources of language actually supports forum interactions that also have desirable characteristics. For example, Table 3 shows that after User 231 noticed that the teacher had made a desirable action, he reflected on why that action was good and even built on that reflection by proposing a useful teaching action to improve what the teacher had done. User 230 then followed up with another viable action of teaching and justified why it would be viable. She also came up with an evaluation of an undesirable action by the teacher and considered an alternative action to correct it.

It would be important to examine whether the posts elicited by posts that contain evaluation also have desirable qualities. Does the likelihood for a post to include reflective comments or alternative teaching actions increase when participants are replying to a post that contains evaluation, compared with when they are replying to a post that does not contain

evaluation? In a preliminary study (Chieu & Herbst, in press) that used a similar analysis method, we found that a forum post had 80.4% chance of including reflection if it replied to a post that did contain evaluation, but only 58.7% chance of including reflection if it replied to a post that did not contain evaluation (effect size or odds ratio = 2.88, $p < .001$). Similarly, a forum post had 58.1% chance of including alternativity if it replied to a post that had evaluation markers, but only 38.5% chance of including alternativity if it replied to a post that did not have evaluation markers (effect size or odds ratio = 2.21, $p < .01$). This finding strengthens the correlations between evaluation and reflection and between evaluation and alternativity through the entire threads of discussion or through interactions among the participants.

The results presented in this paper must be interpreted with caution due to the correlational nature of the claims. Although there exist strong correlations between evaluation and reflection and between evaluation and alternativity, this does not necessarily mean that evaluative comments would lead to reflective comments or proposal of alternative teaching actions. This kind of correlation study is still useful, however, because it provides a valuable foundation for more rigorous research in the future. An experimental design (e.g., the use of control and study groups and random assignment of participants to conditions) with a larger sample size would enable investigations of the effect of evaluative comments on qualities of teacher conversations and comparisons between the use of video records of practice and the use of animated classroom episodes or between face-to-face discussions and online discussions.

Conclusion

In this paper, we have presented strong evidence for correlations between evaluating features of classroom interactions, as indicators, and the probability of making reflective comments and

proposing alternative teaching actions. Indeed, across eight online sessions in a teacher education class both preservice and novice teachers frequently evaluated the teaching practice in animated classroom stories that were embedded into a forum discussion space. Furthermore, the more they were active in evaluating the teaching in the embedded artifacts the more they created, shared, and discussed alternative actions of teaching and the more they reflected on the instructional practice. Those characteristics, reflecting on practice and considering alternatives, have been and continue to be crucial in teacher development, hence it seems important to look for ways to promote them (e.g., Berliner, 1994; Chieu, et al., 2011; Rich & Hannafin, 2009; Schön, 1983; van Es & Sherin, 2008). The data presented shows that evaluative comments on representations of teaching were more likely to include reflection and alternativity than non-evaluative comments. The degree to which such evaluative comments also improve participants' actual teaching is an important question for future research. However, given that reflective practice was found here to be more prevalent when evaluative comments were made, and such reflection has consistently been linked to improved teaching, we infer that such a relationship is likely to be observed.

In producing that finding, we have illustrated how to use Hierarchical Generalized Linear Model (HGLM, a particular form of multilevel modeling, see Hox, 2002; Raudenbush & Bryk, 2002), to examine correlations between different variables of interest (e.g., between *evaluation* and *reflection*) and interactions between different levels of data (e.g., forum posts nested in participants). Research of online conversations or group work has not yet given attention to the nesting of forum posts in individuals, though studies have considered that participants are nested in groups (Cress, 2008; De Wever, Van Keer, Schellens, & Valcke, 2007).

Another key element of this paper is the use of systemic functional linguistics (SFL) to operationalize constructs that are desirable to observe in text-based exchanges among teachers in online discussions. We agree with a number of researchers (e.g., Martin, 2001; Schleppegrell, 2012a, 2012b) that SFL provides a useful means to analyze discourse because it is grounded in a theory of language that simultaneously accounts for the content, the context, and the construction of a discourse.

Our finding can inform the design of facilitation guides for online forum discussions of representations of teaching (Nachlieli, 2011). While scholars have cautioned against promoting evaluation in those discussions to encourage sensitivity, the evidence presented suggests that discouraging participants from making evaluations of the teaching observed might undermine the usage of representations of teaching to promote reflection and alternativity. Instead, if there is concern that participants might shy away from making evaluative comments about colleagues whose teaching has been captured on video (Jacobs, Borko, & Koellner, 2009), developers might consider translating those video records into animations of cartoon characters as a possible way of representing such teaching practice without carrying too much attention to the individual practitioner, so as to have the chance to engage participants in making evaluative comments. Along those lines, promising findings of an earlier study (Kosko & Herbst, 2012; Herbst & Kosko, in press) suggested that animations promoted more uses of modality of the obligation or normativity type (e.g., “the teacher should ...”) than did videos; obligation or normativity is one way in which appraisal is realized. Further research would be needed to investigate the value of animations in terms of supporting evaluation in a safe environment though, particularly comparing online and face-to-face groups.

In terms of a contribution to theory, we argue that it is reasonable that when practitioners get engrossed in a conversation about such a complex, relational practice as teaching, they will engage in it not only intellectually, as an analyst would, but also emotionally, as possible participants of the scenario being discussed. Evaluation, being a function of language with which speakers relate to each other (Martin & White, 2005), could thus be a feature that indicates more rather than less of such engrossment, which the communications literature calls *social presence* (see also Lombard & Ditton, 1997; Oztok & Brett, 2011; Picciano, 2002). We suggest that the ways we have used to estimate reflection, alternativity, and evaluation, as well as the correlations found in this study are important steps in developing ways of estimating the telepresence and social presence of participants from direct observation of their interactions in a discussion forum about teaching.

Acknowledgments

The work reported in this paper is supported by NSF grants ESI-0353285 and DRL-0918425 to Patricio Herbst. Opinions expressed here are the sole responsibility of the authors and do not necessarily reflect the views of the Foundation. We would like to thank the development team of ATutor (<http://www.atutor.ca>), an open-source learning management system on which we have been able to build *LessonSketch*.

References

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, CA: Jossey Bass.

- Barab, S., Kling, R., & Gray, J. (2004). *Designing for virtual communities in the service of learning*. Cambridge, MA: Cambridge University Press.
- Barnett, B. G. (1987). Peer assisted leadership: peer observation and feedback as catalysts for professional growth. In J. Murphy and P. Hallinger (Eds.), *Approaching Administrative Training in Education* (pp. 131-149). Albany, NY: State University of New York Press.
- Berliner, D. (1994). Expertise: The wonder of exemplary performances. In J. Mangieri & C.C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161-186). Ft. Worth, TX: Harcourt Brace College.
- Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3-23.
- Chieu, V.M. & Herbst, P. (in press). Investigating the quality of discussion threads in animation-supported online conversations. To appear in the *Proceedings of E-Learn 2014: World Conference on E-Learning*, New Orleans, LA.
- Chieu, V.M., Aaron, W., & Herbst, P.G. (2013). *Impact of critical events in an animated classroom story on teacher learners' online comments*. Paper presented at the Annual Meeting of AERA, San Francisco, CA.
- Chieu, V.M., & Herbst, P. (2012). LessonSketch: A rich-media scenario-based Learning environment for teacher development. *Proceedings of Society for Information Technology & Teacher Education International Conference 2012*. Austin, TX: AACE.
- Chieu, V.M., Herbst, P., & Weiss, M. (2011). Effect of an animated classroom story embedded in online discussion on helping mathematics teachers learn to notice. *Journal of the Learning Sciences*, 20(4), 589-624.

- Chieu, V.M., & Herbst, P.G. (2011). Supporting mathematics teachers' online discussion with the use of animated classroom stories as reference point. *In Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies* (pp. 479-481). DOI 10.1109/ICALT.2011.149. Washington, DC: IEEE Computer Society.
- Colasante, M. (2011). Using video annotation to reflect on and evaluate physical education pre-service teaching practice. *Australasian Journal of Educational Technology*, 27(1), 66-88.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research — An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 69-84.
- Edens, K. M. (2000). Promoting communication, inquiry and reflection in an early practicum experience via an on-line discussion group. *Action in Teacher Education*, 22(2A), 14-23.
- Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Meittinen, & R. Punamaki (Eds.), *Perspectives on activity theory* (pp. 19-38). New York, NY: Cambridge University Press.
- Fishman, B. J., & Davis, E. A. (2006). Teacher learning research and the learning sciences. In R. Keith Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 535-550). Cambridge, MA: Cambridge University Press.
- Goldsmith, L.T., & Seago, N. (2011). Using classroom artifacts to focus teachers' noticing: Affordances and opportunities. In M.G. Sherin, V.R. Jacobs, & R.A. Philipp, *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 169-187). New York, NY: Routledge.
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055-2100.

- Hadden, D. S., & Pianta, R. C. (2006). Clinical consultation with teachers for improved preschool literacy instruction. In L. M. Justice (Ed.), *Clinical approaches to emergent literacy instruction* (pp. 99-124). San Diego, CA: Plural Publishing.
- Halliday, M.A.K., & Matthiessen, C. (2004). *An introduction to functional grammar*. London, England: Arnold-The Hodder Headline Group.
- Herbst, P., Aaron, W., & Chieu, V.M. (2013). LessonSketch: An environment for teachers to share and examine mathematical practices. In D. Polly (ed.), *Common Core Mathematics Standards and Implementing Digital Technologies* (pp. 281-294). IGI Global.
- Herbst, P., Aaron, W., Dimmel, J., & Erickson, A. (2013). *Expanding students' involvement in proof problems: Are geometry teachers willing to depart from the norm?* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco. Available at Deep Blue at the University of Michigan. <http://hdl.handle.net/2027.42/97425>
- Herbst, P., & Chazan, D. (2003). Exploring the practical rationality of mathematics teaching through conversations about videotaped episodes. *For the Learning of Mathematics*, 23(1), 2-14.
- Herbst, P., & Chazan, D. (2006). *Producing a viable story of geometry instruction: What kind of representation calls forth teachers' practical rationality?* Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Merida, Mexico.
- Herbst, P., Chazan, D., Chen, C., Chieu, V.M., & Weiss, M. (2011). Using comics-based representations of teaching, and technology, to bring practice to teacher education courses. *ZDM—The International Journal of Mathematics Education*, 43(1), 91-103.

- Herbst, P. & Kosko, K.W. (in press). Using representations of practice to elicit mathematics teachers' tacit knowledge of practice: a comparison of responses to animations and videos. *Journal of Mathematics Teacher Education*, Online First DOI 10.1007/s10857-013-9267-y.
- Herbst, P., & Miyakawa, T. (2008). When, how, and why prove theorems? A methodology for studying the perspective of geometry teachers. *ZDM—The International Journal of Mathematics Education*, 40, 469-486.
- Herbst, P., Nachlieli, T., & Chazan, D. (2011). Studying the practical rationality of mathematics teaching: What goes into “installing” a theorem in geometry? *Cognition and Instruction*, 29(2), 218-255.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Jacobs, J., Borko, H., & Koellner, K. (2009). The power of video as a tool for professional development and research: Examples from the Problem-Solving Cycle. In Janik, T. & Seidel, T. (Eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom* (pp. 259-273). Munster, Germany: Waxmann.
- Joyce, B. & Showers, B. (1980). The coaching of teaching. *Educational Leadership*, 40(1), 1980, 4-10.
- Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design*. Cambridge, MA: The MIT Press.
- Kelchtermans, G. & Vandenberghe, R. (1994). Teachers' professional development: A biographical perspective. *Journal of Curriculum Studies*, 26, 45-62.

- Knight, J. (2006). Instructional coaching: Eight factors for realizing better classroom teaching through support, feedback and intensive, individualized professional learning. *The School Administrator*, 63(4), 36-40.
- Kosko, K. & Herbst, P. (2012). A deeper look at how teachers say what they say: A quantitative modality analysis of teacher-to-teacher talk. *Teaching and Teacher Education*, 28, 589-598.
- Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education*, 61(1-2), 21-34.
- Lampert, M., & Ball, D. L. (1998). *Teaching, multimedia, and mathematics: Investigations of real practice*. New York, NY: Teachers' College Press.
- LeFevre, D. M. (2004). Designing for teacher learning: Video-based curriculum design. In J. Brophy (Ed.), *Using Video in Teacher Education (Advances in Research on Teaching, Vol. 10, pp. 235-258)*. Oxford, UK: Elsevier Ltd.
- Leontiev, A. (1978). *Activity, consciousness, and personality*. Englewood Cliffs, NJ: Prentice Hall.
- Liu, X., Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S. (2007). *Optimal design for multi-level and longitudinal research: Version 1.83* [Statistical software]. Scientific Software International, Inc.
- Lombard, M. & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2).
- Males, L.M., Otten, S. & Herbel-Eisenmann, B.A. (2010). Challenges of critical collegueship: Examining and reflecting on study group interactions. *Journal of Mathematics Teacher Education*, 13, 459-471.

- Martin, J.R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.). London, the United Kingdom: Continuum.
- Martin, J.R., & White, P.R.R. (2005). *The language of evaluation: Appraisal in English*. New York, NY: Palgrave MacMillan.
- Martin, J. R. (2000). Beyond exchange: APPRAISAL systems in English. *Evaluation in text*, 142-175.
- Martin, J. R. (2001). Cohesion and texture. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 35-53). Malden, MA: Blackwell.
- Moore-Russo, D., & Viglietti, J.M. (2011). Teachers' reactions to animations as representations of geometry instruction. *ZDM—The International Journal of Mathematics Education*, 43(1), 161-173, DOI 10.1007/s11858-010-0293-2.
- Moore-Russo, D. & Wilsey, J. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflections on representations of teaching. *Teaching and Teacher Education*, 37, 76-90.
- Moreno, R., & Ortegado-Layne, L. (2008). Do classroom exemplars promote the application of principles in teacher education? A comparison of videos, animations and narratives. *Educational Technology Research and Development*, 56(4), 449-465.
- Nachlieli, T. (2011). Co-facilitation of study groups around animated scenes: the discourse of a moderator and a researcher. *ZDM*, 43(1), 53-64.
- Nachlieli, T., & Herbst, P., with González, G. (2009). Seeing a colleague encourage a student to make an assumption while proving: What teachers put to play in casting an episode of geometry instruction. *Journal for Research in Mathematics Education*, 40(4), 427-459.

- Neale, D. C., Carroll, J. M., & Rosson, M. B. (2004). Evaluating computer-supported cooperative work: Models and frameworks. In J. Herbsleb & G. Olson (Eds.), *Proceedings of Computer Supported Cooperative Work 2004* (pp. 112-121). New York, NY: ACM Press.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*(4), 317-328.
- Oztok, M., & Brett, C. (2011). Social presence and online learning: A review of research. *The Journal of Distance Education, 25*(3).
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K. & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431-451.
- Picciano, A. G. (2002). Beyond student perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous learning networks, 6*(1), 21-40.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Scientific Software International, Inc.
- Rich, P.J., & Hannafin, M. (2009). Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education, 60*(1), 52-67.
- Schiffrin, D., Tannen, D., & Hamilton, E. (2003). *The handbook of discourse analysis*. Malden, MA: Wiley-Blackwell.
- Schleppegrell, M.J. (2012a). Linguistic tools for exploring issues of equity. In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in Discourse for Mathematics Education: Theories, Practices, and Policies* (pp. 109-124). New York, NY: Springer.

- Schleppegrell, M.J. (2012b). Systemic Functional Linguistics: Exploring meaning in language. In J.P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 21-34). New York, NY: Routledge.
- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Scott, A.L. (2005). Pre-service teachers' experiences and the influences on their intentions for teaching primary school mathematics. *Mathematics Education Research Journal*, 17(3), 62-90.
- Seago, N. (2004). Using video as an object of inquiry for mathematics teaching and learning. In J. Brophy (Ed). *Using video in teacher education, Advances in Research on Teaching, Vol. 10*, pp. 259-286. Amsterdam: Elsevier.
- Sherin, M.G., Jacobs, V.R., & Philipp, R.A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York, NY: Routledge.
- Sherin, B.L., Sherin, M.G., Colestock, A.A., Russ, R.S., Luna, M.J., Mulligan, M., & Walkoe, J. (2010). Using digital video to investigate teachers' in-the-moment noticing. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 2, Short Papers, Symposia, and Selected Abstracts* (pp. 179-186). Chicago, IL: International Society of the Learning Sciences.
- Sim, J., & Wright, C.C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257-268.
- Star, J.R., & Strickland, S.K. (2008). Learning to observe: Using video to improve preservice teachers' ability to notice. *Journal of Mathematics Teacher Education*, 11(2), 107-125.

- Stiler, G. M., & Philleo T. (2003). Blogging and blogspots: An alternative format for encouraging reflective practice among preservice teachers. *Education, 123*(4), 789-797.
- Tettegah, S., Whang, E., Taylor, K., & Cash, T.J. (2008). Narratives, virtual environments and identity semiotics: An exploration of pre-service teacher's cognitions. *Journal of E-Learning, 5*(1), 103-127.
- Tettegah, S. (2005). Technology, narratives, vignettes, and the Intercultural & Cross Cultural Teaching Portal. *Urban Education, 40*(4), 268-293.
- Tschannen-Moran, B. & Tschannen-Moran, M. (2011). The Coach and the evaluator. *Educational Leadership, 69*(2), 10-16.
- 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.
- Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wilkins, J. L. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139-164.
- Wise, A.F., Padmanabhan, P., & Duffy, T.M. (2009). Connecting online learners with diverse local practices: The design of effective common reference points for conversation. *Distance Education, 30*(3), 317-338.
- Zhang, M., Lundeberg, M.A., Koehler, M.J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education, 27*(2), 454-262.