HOW DO TEACHERS EXPECT STUDENTS TO REPRESENT MATHEMATICAL WORK? A STUDY OF TEACHERS' RECOGNITION OF ROUTINE WAYS THAT PROOFS ARE PRESENTED AND CHECKED IN HIGH SCHOOL GEOMETRY

by

Justin K. Dimmel

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2015

Doctoral Committee:

Professor Patricio G. Herbst, Chair
Professor Hyman Bass
Professor Daniel I. Chazan, University of Maryland
Professor Mary J. Schleppegrell
Professor Edward A. Silver

To the Dimmel and Lilleston families.

ACKNOWLEDGMENTS

I am grateful to the University of Michigan, the School of Education, and the Rackham Graduate School for the opportunity to pursue a doctor of philosophy in mathematics education. I am grateful to the Mathematics Department for the opportunity to pursue a masters degree in mathematics in parallel with my doctoral studies. I was able to spend five years in Ann Arbor because of the generous support of a Rackham Merit Fellowship. At every step along the way to completing my degrees, the institutional support of the University of Michigan helped keep me on course.

I have had the privilege of working closely with Professor Pat Herbst, my advisor and dissertation committee chair, throughout my time at Michigan. I am thankful that I was apprenticed into the practice of mathematics education scholarship through opportunities he provided. I have learned to balance the duty to be self-critical against the drive to be creative by following his example. I thank Professor Herbst for his consistent guidance, support, and challenges to produce high-quality work.

The members of my dissertation committee have been generous with their time and expertise and have pushed me to make my work stronger. I thank Professor Hyman Bass, Professor Daniel Chazan, Professor Mary Schleppegrell, and Professor Edward Silver for their questions, comments, and insights into my work. Their attention helped me to think carefully about the big picture of my study and of education as a field of inquiry.

During my time at Michigan, the GRIP research group (directed by Professor Herbst) has provided a community of colleagues and friends whose support has been instrumental to my scholarly development. I thank Wendy Aaron, Kayin Aaron, Vu Minh Chieu, Annick Rougee, Karl Kosko, Amanda Milewski, Kristi Hanby, Inah Ko, Nicolas Boileau, and Umut Gursel for their parts in creating an environment that is a model of collaborative research in mathematics education. I also wish to recognize the friendship and support of Rohit Setty, Tim Whittemore, Rachel Snider, Elaine Lande, and Anne Cawley. In addition, I want to thank GRIP member Ander Erickson for five years of camaraderie.

I would not be where I am today without my siblings, Kevin and Kaitlyn. My siblings have provided comfort, laughter, and unflinching support throughout my life. Our father, George Dimmel, passed away before he had a chance to see the kind of people his children would grow up to become. As I have watched my siblings switch careers, earn graduate degrees, raise children, and excel in all aspects of grown up life, I know in my heart that our father would be proud.

My mother, Vivian Dimmel, worked her entire life to make sure that her children would have the opportunities to be whatever they wanted to be. She encouraged me to invest in my education and was thrilled to be a Michigan mom. She passed away this fall, but not before I had the chance to tell her that I had submitted my dissertation. For the years of her life that she dedicated to raising, nurturing, and loving me, I thank my mom.

Finally, I want to thank my wife, Amanda, for helping get me to the finish line. She has been patient, supportive, and a constant source of comfort throughout our years of partnership, but especially during the final months of writing my dissertation. Thinking of the life that

Amanda and I are building together helped me find motivation during the most challenging times. Special thanks, also, to Wookie (our pug).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation is an investigation of how teachers expect students to represent mathematical work. The overall goal of the study is to identify routine ways that students communicate in mathematics classrooms and to determine whether mathematics teachers recognize these routines. The instructional setting of the study is US high school geometry. The research reported here looks specifically at how students are expected to communicate when doing proofs.

The study consists of two parts. The first part of the study examined video episodes of geometry classrooms to identify how students use different modes of communication when presenting and checking proofs in geometry classrooms. From the analysis of video episodes, I ground hypotheses of routine ways in which students use communication modalities; I call them *semiotic norms.*

The second part of the study is an experiment that uses representations of geometry instruction to investigate the extent to which secondary teachers recognize specific semiotic norms that I call *details* and *sequence*. The *details* norm describes what students are expected to include in the written statements of a proof. The *sequence* norm describes the expected order of events contributing to the writing and reading of proofs that students present proofs to the class.

The experiment conducted for the second part of the study used storyboards that represent episodes of geometry classrooms as probes for a multimedia questionnaire. Participants

in the experiment viewed storyboards that represented teachers breaching or complying with the hypothesized norms. Seventy-three high school mathematics teachers from schools within a 60 mile radius of Midwestern University completed the questionnaires. The results of the experiment indicate that secondary mathematics teachers recognize that the *details* and *sequence* norms describe routine communication practices of the activity of doing proofs in geometry.

The work reported here identifies communication practices that students use in geometry classrooms when doing proofs. By describing these practices, the research reported in this dissertation contributes subject-specific knowledge of what routinely happens in mathematics classrooms. Knowledge of the routine ways that students communicate is valuable because it provides a foundation for developing discipline-specific communication skills in mathematics classrooms. In turn these inform our understanding of literacy practices in the mathematics classroom.

CHAPTER 1:

INTRODUCTION

The purpose of this study of mathematics teaching is to describe activities that take place in mathematics classrooms. As a descriptive study, the research reported here is in the same vein as the TIMSS video study of German, US, and Japanese mathematics classrooms (Stigler et al. 1999). A result of the TIMSS study was the identification of *teaching scripts*—signature ways that the classroom activity in different countries unfolds. The TIMSS study used analyses of video episodes of mathematics classrooms as the basis for identifying teaching scripts. The force of the findings of the TIMSS study stems from the fact that the researchers were able to record video from a nationally representative sample of mathematics classrooms in each country (Stigler et al., 1999).

The teaching scripts identified in the TIMSS study are coarse accounts of how mathematics instruction typically unfolds. For example, part of the US teaching script identified in the TIMSS study is: “The mathematics teacher sets the goal for the lesson as the acquisition of a skill or procedure for solving a mathematical problem” (Stigler et al. p. 134). The coarseness of the scripts is a necessary consequence of the design and scale of the study. An enduring takeaway of the TIMSS study is that the activities that take place in different mathematics classrooms have commonalities that can be investigated by research. The TIMSS study provides a base of empirical evidence to support the notion that describing what happens in mathematics classrooms is a means of generating knowledge about mathematics education.

Like the TIMSS study, the goal of the research conducted for this dissertation is to identify regularities of mathematics classrooms. Whereas the regularities reported in TIMSS are general and based on observations from a nationally representative sample of classroom episodes, the regularities reported in this dissertation are specific and based on a mixed methods design that used records of practice cyclically (Jacobs, Kawanaka, and Stigler, 1999). The cyclical use of records of practice involved analyzing a small sample of video records of mathematics classrooms to conjecture regularities of classroom activity and then experimentally testing whether mathematics teachers recognized those regularities. The mixed methods design provided a means to investigate regularities that are mathematically specific.

I use *regularity of classroom activity* as a generic way of referring to the routine, usual, or typical things that happen in mathematics classrooms. An example of a regularity of instruction that is not unique to mathematics classrooms is the physical arrangement of the desks in the room into rows and columns. A related regularity is the practice of assigning students seats in this row/column grid alphabetically by last name. A third regularity related to the arrangement of classroom space is for the teacher's desk to be positioned at the front of the room, offset from the center of the central classroom board.

The sense in which such ways of arranging the classroom space are routine is shared and social. By this I mean that the claim that such ways of arranging the furniture in a classroom are routine does not depend on the number of US classrooms that use this arrangement of desks. By stating that an alphabetized, row/column arrangement of classroom desks is a routine occurrence in classrooms, I am not claiming that this is an empirical fact of classrooms. The actual percentage of US classrooms in public schools that arrange furniture in this way does not change the fact that such arrangements of furniture would be recognized by people familiar with classrooms as "nothing out of the ordinary". Even teachers that use different arrangements of desks in their own classrooms would likely recognize[1] that the row/column layout is a typical or default option for arranging the furniture in classrooms.

In broad terms, the regularities investigated for this study concern how teachers expect students to represent mathematical work. I focus on student work because of the central role that student work plays in the activity that unfolds in mathematics classrooms. It is a reality of schooling that the mathematical knowledge and practices that students are able to learn are mediated through the mathematical work students are required to do (Doyle, 1988; Herbst, 2006; Kilpatrick, 2001). This means that researchers can investigate the opportunities for learning mathematics that are typically available to students in classrooms by describing classroom interactions that take place around student work..

An objective of this study is to describe how the work that students produce in mathematics classrooms creates opportunities for students to develop their mathematical communication skills. The communicative turn in mathematics education research has produced a bevy of scholarship from a range of theoretical perspectives about the role that language, discourse, and mathematical communication play in the teaching and leaning of mathematics (Morgan, Craig, Schuette, & Wagner, 2014; Walshaw & Anthony, 2008). The underlying issues about the nature of the relationship between knowledge and communication in mathematics have vexed philosophers for millennia and are not likely to be decisively settled by mathematics educators anytime soon.[2] Though I do not wish to argue for one position, I believe investigating how students are taught to communicate about mathematics is productive because the ability to

[1] What it means to recognize a regular aspect of a social activity is discussed below.
[2] Sierpinska (2005) and Morgan et al. (2014) offer broad assessments of some of these issues in terms of the current state of research into mathematical discourse

observe, describe, or measure mathematical knowledge depends on systems of representation through which such knowledge can be communicated[3]. As a researcher who wants to understand the activity that takes place in mathematics classrooms, I have found that communicative acts are manifestly observable in ways that internal acts of thinking are not.

The study of communication[4] in mathematics classrooms is not an end in itself (Sierpinska, 2005). By this I mean to draw a distinction between communication, in general, and the discipline-specific communication practices that mathematics students might be expected to develop in mathematics classrooms (Lemke, 2013; Fang, 2012). Mathematical experts communicate in discipline-specific ways that are linked to particular kinds of mathematical activities. As a example, the activity of preparing a proof to be published in a peer-reviewed journal calls for different kinds of communication than the activity of presenting a proof to an undergraduate classroom. In the case of the former, part of the communication involves knowing what not to write in a proof (Davis, 1972) (i.e., what aspects of a mathematical argument can be elided because they are generally known or follow immediately from known results?), for the latter, part of the communication involves situating the result that is to be shown in an appropriate theoretical context (Greiffenhagen, 2014) (i.e., providing the background on how the result fits with related mathematical concepts). There are still more fine-grained differences that could be surfaced by looking closely at practices for communication in specific fields of mathematics.

The communication skills used by disciplinary experts have traditionally been thought to be gradually and tacitly developed by novices as the novices are apprenticed into the field

---

[3] One could draw an analogy to the classic riddle about a tree falling in the woods: If a conjecture is decided in a mathematician's mind…

[4] I use *communication* to be a general term that encompasses classroom discourse that uses linguistic and non-linguistic (gestures, pictures, tools) semiotic resources.

(Lemke, 2013; Thurston, 1994). But recent work in analyzing mathematical communication suggests that discipline-specific ways of communicating are practices that can be described and taught (Fang, 2012; O'Halloran, 2005; O'Halloran, 2011; Yore, Pimm & Tuan, 2007). Describing classroom activity during which students have the opportunity to use discipline-specific communication skills is a step in the direction of devising teaching strategies that would help students develop these skills directly.

One classroom setting where disciplinary communication practices have the opportunity to be foregrounded is in high school geometry[5]. The geometry classroom has historically been the principal instructional setting in which students are introduced to mathematical proof (González & Herbst, 2006; Herbst, 2002; Knuth, 2002). Though the instructional activity of doing proofs in geometry has been criticized by mathematics educators for being a misrepresentation of the work of proving in mathematics (Schoenfeld, 1988; Martin & Harel, 1989; Lockhart, 2009), it endures as an instructional setting where students are introduced to the notion that there is such a thing as mathematical proof. Investigating how teachers expect students to represent mathematical proofs in geometry classrooms is a means of analyzing the opportunities for learning disciplinary communication practices that are generally available in school mathematics.

Proving in mathematics is an activity that has been likened to an orchestra's performance of a symphony (Herbst & Balacheff, 2009). In this analogy, what teachers commonly refer to as "formal (written) proofs"—such as those found in mathematics textbooks—are the scores (i.e., sheet music) that represent a (mostly) invariant description of the music the orchestra plays, while the activity of proving is the actual concert. I draw this distinction between performances

[5] Throughout this study, "geometry classrooms" will be used as a short hand to refer to the proof-based secondary geometry course that is taught to US students.

of proofs and written descriptions of proofs to foreground the need to describe a wider range of semiotic resources, such as what a prover says, draws, or gestures, when investigating proof in classrooms. This semiotic widening is warranted because "language is at the core of argumentation and proof in general, but especially in mathematics" (Balacheff, 2008, p. 502), and the language of a proof involves more than just writing (Livingston, 1999). The aim of the study reported here is to describe the work that students are accountable for producing when doing proofs in geometry.

Providing such a description is important for the field of mathematics education and for the teaching of mathematics. For the field, identifying mathematically specific regularities of classroom activity that are recognizable to teachers in general advances the work of describing units of classroom activity that can be scientifically studied (Arsac, Balacheff, & Mante, 1992; Herbst 2002a, 2006). The research reported here furthers these efforts by showing that regularities of the way that proofs are communicated can be investigated through an experimental study. For the teaching of mathematics, uncovering the ways that teachers expect students to communicate when doing proofs provides diagnostic information that could identify opportunities for students to develop disciplinary literacy in mathematics classrooms.

The description of doing proofs reported in this study investigated two research questions:

1. How do students use semiotic resources when doing proofs in geometry?
2. To what extent do secondary teachers recognize routine ways semiotic resources are used when doing proofs?

I investigate these questions in a two-part study of the instructional situation of doing proofs in geometry (Herbst & Brach, 2006). The first part of the study is a description of episodes of geometry classrooms in which: (1) students present proofs during geometry class, and (2) teachers check completed proofs during geometry class. I describe these activities as *presenting* and *checking* proofs throughout this study. The goal of the first part of the study was to identity the ways that teachers expect students to communicate when doing proofs in geometry.

The second part of the study is a *virtual breaching experiment with control* that uses representations of geometry classrooms to examine secondary teachers' reactions to episodes of instruction that vary in planned, controlled ways. The episodes used as probes for this second part of the study are instances of students presenting proofs or teachers checking a proof with the class. The goal of the second part of the study is to determine the extent to which teachers recognize the routine ways of communicating about proofs that are hypothesized during the first part of the study.

Descriptions of *presenting* and *checking* proofs

A routine expression of the communication practices of expert mathematicians is *chalk talk*, i.e., the practice of generating a proof at a black board (Artemeva & Fox, 2011). During chalk talk, a mathematician coordinates speaking, writing (word, notation), drawing (diagrams or other mathematical pictures), and gesturing to develop a mathematical result—usually a proof—in real time (Artemeva & Fox, 2011; Greiffenhagen, 2014; Núñez, 2009). The communication practices associated with *chalk talk* are indicators of the mathematical proficiency (Kilpatrick, 2001) of mathematicians (Artemeva & Fox, 2011). The presentation of a proof in a geometry classroom is an opportunity for students to engage in an approximation of *chalk talk.*

The checking of a proof by the teacher is significant because it is a means through which teachers help students develop their sense of when a proof is complete. The activity of checking a proof is a step-by-step review of each line of a proof, the end goal of which is to eliminate logical gaps between steps. The video episodes of teachers checking the details of proofs as a whole-class activity thus provides a view into how teachers help students manage the exchange of the work they do (i.e., the written statements and reasons they produce) for the knowledge at stake in the work.

The aim of describing how proofs are presented and checked in geometry classrooms is to uncover the routine ways semiotic resources are used when doing proofs in geometry. I refer to the routine ways of using semiotic resources to present and check proofs as *semiotic norms.* The primary outcome of the first part of the study is a list of semiotic norms of doing proofs in geometry that I hypothesize are recognizable to secondary mathematics teachers. The goal of the second part of the study is to test whether or not teachers recognize these semiotic norms.

A breaching experiment with treatment and control conditions

I use an experimental study to examine the extent to which secondary mathematics teachers recognize semiotic norms of doing proofs in geometry. The ethnomethodological notion of a *breaching experiment* (Garfinkel, 1963) holds that there are norms that structure routine social interactions, and, for the most part, these norms are tacit; i.e., a norm is transparent or invisible unless it is breached. But when such norms are breached, people notice the breach and make some maneuvers to *repair* the (no longer normal) situation. A repair of the situation is not necessarily a repair in the sense of a fix, but rather is a signal that the situation is not normal.

Granting that *situations are not remarked on* (Q) whenever *they conform to their norms* (P), the logic of a breaching experiment is a form of *modus tollens*: if a situation is remarked on as being somehow strange (~Q), then a norm of the situation has been breached (~P). The technique of the breaching experiment thus provides a means of empirically testing whether a feature of a situation that is hypothesized to be a norm is in fact recognized as such by people who typically participate in that situation. For this study, I use an experimental variant of the classic breaching experiment technique to test for recognition of semiotic norms.

Mathematics education researchers (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009; Herbst et al., 2013) have used virtual breaching experiments to gauge whether mathematics teachers recognize breaches of (hypothesized) norms of instructional situations and track how those teachers repair the situation. While initial breaching experiments of this virtual sort were facilitated with video (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009), the breaching experiments described by Herbst and Chazan (2011, 2012) used cartoon representations of classroom episodes to depict breaches of norms. Cartoon depictions of episodes of instruction have been shown to be an effective way of representing classroom scenarios to teachers (Chazan & Herbst, 2012; Chieu, Herbst, & Weiss, 2011; Herbst et al., 2013), and using such depictions has the advantage that they can be customized to represent episodes of instruction that are hypothetical. Researchers have also shown that cartoon representations of classroom episodes function comparably to classroom videos when used to prompt teachers for commentary (Herbst & Kosko, 2014).

Overview of the dissertation

The research presented in this dissertation is organized into seven chapters. Chapter 2 provides a review of literature on proof and a theoretical framework for investigating doing proofs in geometry as an instructional situation. Chapter 3 reports the descriptions of episodes of students presenting proofs and teachers checking proofs in geometry classrooms. The methods used to select and describe the episodes are included at the beginning of the chapter. Chapter 4 describes the design of the experiment that was developed to gauge the extent to which teachers recognize semiotic norms. Chapter 5 reports the results of the analysis of the closed-ended questions participants were asked during the experiment. Chapter 6 describes the coding schemes that were developed to analyze responses to the open-ended questions participants were asked during the experiment. The results of the analysis of the open response data are also reported in Chapter 6. The discussion and conclusion are presented in Chapter 7.

CHAPTER 2:

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

In this chapter I review literature and present a theory of instruction that frames the studies described in Chapter 3 and Chapter 4. The literature review begins with a consideration of the instructional and conceptual challenges of teaching and learning proof in mathematics classrooms. Following these considerations, I home in on accounts of disciplinary practices for communicating proofs. The purpose of the literature review is to motivate the descriptions of instances of *presenting* and *checking* proofs provided in Chapter 3.

The theoretical framework conceptualizes instruction as a social system that is achieved by teachers and students playing roles in classrooms. For this I draw on the theory of instructional exchanges (Herbst, 2006) that describes the activity that takes place in mathematics classrooms as transactions in which teachers trade student work for knowledge claims.. I use this model of instruction to warrant the focus on describing teacher and student interactions around representations of student work that are reported in Chapter 3. I conclude the chapter with a summary and a preview of the studies described in Chapters 3 and 4.

Instructional and conceptual challenges of teaching proof in schools

A primary reason that proof is taught in schools is that “proof is intimately connected to the construction of mathematical ideas” (Herbst, 2002, p. 283). Despite intractable epistemic and ontological debates about the nature of mathematics (Rota, 1997), the nature of proof (Balacheff, 1991), and the nature of mathematical ideas (Hanna, 2000), there is a general consensus that mathematical proof is a means through which mathematical knowledge grows (Knuth, 2002; Herbst, 2002, 2002a). Proof and proving are taught in schools "because they are fundamental to doing and knowing mathematics; they are the basis of mathematical understanding and essential in developing, establishing, and communicating mathematical knowledge" (Stylianides, 2007, p. 289). Developing, establishing, and communicating mathematical knowledge are disciplinary practices that learning proof in schools could create occasions for students to approximate.

I say *could* create because whether proof as it unfolds as an activity in schools serves as a viable representation of disciplinary practices of proof has been a fundamental question for research on proof and proving (Herbst, 2002). The challenges facing teachers when teaching proof in schools are salient. For example, managing the tension between proof as a disciplinary practice and proof as a didactic activity can put the teacher in a double bind (Herbst, 2002): On the one hand, posing proof problems that create opportunities for students to engage in disciplinary proving practices (e.g., formulating a conjecture; hypothesizing its truth value) can create a task that is too challenging (or too alien; see Herbst, 2006) for students to complete on their own. A teacher in such a situation might default to providing extensive help to the students (Herbst, 2006). On the other hand, posing proof problems that students could be expected to do—i.e., standard proof exercises in geometry textbooks, with given information, a statement to prove, and a diagram—diminishes the work required of the students to complete the task (Schoenfeld, 1988)—thus distancing the proofs done by students from disciplinary proof practices (Herbst, 2002). In addition to the didactic analysis provided by Herbst (2002), the challenges of teaching proof in classrooms have also been studied from the perspective of teachers' and students' conceptions of what proof is. Several papers from the literature on teachers' and students' conceptions of proofs encapsulate these challenges.

Martin and Harel (1989) investigated the proof frames of pre-service elementary school teachers. They found that “many students accepted inductive arguments as proofs of mathematical statements” (Martin & Harel, 1989, p. 48) Inductive arguments—i.e., those drawing a general conclusion from a consideration of finite cases, not to be mistaken for arguments that hinge on the strong or weak principles of mathematical induction—are examples of *empirical* proof, a specter that has garnered considerable attention in the research about students’ and teachers’ difficulties in understanding mathematical proof.

Chazan (1993) identifies two general categories of misconceptions about the nature of proof and evidence in geometry that are intertwined with the status of empirical evidence in mathematical argument. One misconception is that empirical evidence—e.g., measures of the parts of specific geometric figures—can be used to establish the truth value of general mathematical claims—viz. the diagonals of a rectangle are congruent. This misconception treats specific empirical facts as sufficient to establish universal claims. A complementary misconception is that a deductive proof only applies to the specific example that is referenced in the diagram that accompanies the proof. This misconception treats the generic mathematical figure as if it were just a special case.

Chazan (1993) designed a study to uncover the justifications that students use for their beliefs about the roles played by different kinds of evidence when investigating the truth of a mathematical claim. The study relied on interviews of 17 geometry students, during which the students were shown two example proofs: an exemplar of a deductive proof (in two-column format) and an exemplar of an empirical argument (based on measurements of several cases). Chazan (1993) found that there was evidence in the students he interviewed for the two sets of beliefs about proof: empirical evidence is proof/deductive proof has a limited range. Chazan

(1993) also found that "students had good reason to believe that evidence is proof, especially in the realm of triangles" (p. 383), for instance by arguing for the need to try a variety of examples to ensure that the claim was not being based on a special case.

Harel and Sowder (1998) further investigated students' perspectives on the nature of the evidence that is necessary to establish the truth of a mathematical claim in their exploratory study of student proof schemes. One legacy of their research is the finding that the "*empirical proof scheme* is pervasive among school students" learning mathematics at all levels and all ranges of abilities (Stylianides & Stylianides, 2009, p. 315). An empirical proof scheme is one in which empirical evidence—e.g., measurements of specific objects, a set of specific cases—is used to warrant general mathematical conclusions. Beyond the sway that empirical evidence has for student evaluations of mathematical arguments, studies of teachers' conceptions of proof indicate that empirical evidence also plays a role in how teachers evaluate mathematical arguments.

Knuth (2002, 2002a) used semi-structured interviews to explore secondary teachers' conceptions of proofs. Teachers were shown various written representations of proofs and asked to discuss the extent to which they found the proofs convincing. The arguments that the teachers were shown "varied in terms of their validity as proofs" and were chosen so that the underlying mathematical concepts at stake in the proofs were not difficult (Knuth, 2002, p. 383). The goal was to focus the review of each proof "on the argument presented rather than on trying to understand the mathematics needed to produce the argument" (Knuth, 2002, p. 383). Teachers were shown a mixture of valid and invalid proofs as well as arguments that "varied in terms of the approach used in constructing them (e.g., algebraic, proof by induction) and their explanatory nature (i.e., more explanatory or less explanatory)" (Knuth, 2002, p. 391). Five of the 16 teachers in the study identified an empirical argument as valid mathematical proof. Beyond the fact that some teachers identified empirical arguments as proof, “many teachers” in the study used empirical evidence to test a proof’s conclusion (Knuth, 2002, p. 401).

Another legacy of the research into proof schemes (Harel & Sowder, 1998) is a preoccupation in the literature with delineating grounds on which an argument should or should not *count* as a proof. Dreyfus (1999) frames the issue as a matter of identifying the warrants teachers use to accept (or not accept) a student’s explanation as mathematically valid. Stylianides and Stylianides (2009) offer the tautology that proofs that count are those that use valid modes of argumentation and proofs that do not (or should not) count are those that use invalid modes of argumentation. Weber (2008) argues that “more research is needed on the practice of and processes involved in determining whether an argument constitutes a valid proof” (p. 432). Stylianides (2007) places the notion of what counts as a proof at the center of the framework he elaborates for investigating proof across grade levels.

Stylianides (2007) offers a conceptualization of proof that is intended to be valid from the elementary grades through secondary school. This conceptualization of proof is first and foremost a definition of proof: “A *proof* is a *mathematical argument,* a connected sequence of assertions for or against a mathematical claim” that uses true, available statements; employs valid modes of argumentation; and is presented using appropriate modes of argument representation (Stylianides, 2007, p. 291). The issues of truth, validity, availability, and appropriateness of the aspects of proof invoked in the definition are relative to mathematical discourse in specific mathematical communities. Thus, the modes of argument representation that are appropriate in a first-grade classroom will be different from those that are appropriate in high school geometry.

Stylianides (2007) used this definition to build an analytic machine that can decide whether or not an argument counts as a proof in a particular classroom setting (Stylianides, 2007, p. 299-300). The procedure he describes uses the definition of proof—in a specific classroom setting—to define local validity criteria an argument would need to exhibit before it would count as proof in that setting. Stylianides (2007) describes the analytic comparison between *base arguments*—i.e., the arguments that students initially formulate—and the definition of proof: "In each episode, I identified the base argument. Then, by using the definition of proof, I Identified what could count as proof in each episode. The comparison between the base argument and proof allowed me to see how close the students' initial proving activity was to the construction of a proof" (p. 313).

The motivation for Stylianides (2007) prescriptive conceptualization of proof is to provide an operational framework for teachers at all levels to guide the mathematical activity of their students toward realizations of the classroom activity of proving that are consistent with disciplinary proof practices. The three components of the definition of proof that Stylianides (2007) proposes also provide a scaffold for teachers to target particular elements of student arguments. From the perspective of describing the activities of the teacher in the classroom, Stylianides's (2007) definition of proof is operational in so far as it allowed him to "satisfactorily explain the proving activity in which the students and teacher engaged" (p. 313).

The literature considered above were (1) studies of the cognitive and psychological aspects of student and teacher knowledge about mathematical proof, or (2) studies of how proofs are negotiated by communities of learners in classrooms. The work conducted for this study contrasts with each of these strands of research. Whereas research that describes how proofs are done have studied individuals (e.g., Harel & Sowder, 1998; Knuth, 2002), my work aims to describe collective classroom practices around proof and proving. Whereas studies of classroom discourse around proof have used prescriptive criteria to evaluate the mathematical quality of the activity that takes place (e.g., Stylianides, 2007; Stylianides & Stylianides, 2009), my work aims to uncover and describe the tacit aspects of classroom practice that make the activity of proving recognizable to those that engage in the activity (Herbst 2002a, Herbst 2006). The next section of the literature review considers studies of how proofs are enacted in classrooms before moving to a consideration of how proofs are communicated.

Studies of enactments of proof in schools

Relatively few studies have investigated enactments of proof in mathematics classrooms (Bieda, 2010; Mariotti, 2006). In a study of the enactment of proof tasks in high school geometry classrooms, Sears Chávez (2014) analyzed the level of cognitive demand of the tasks at various stages in their implementation by teachers. The design of the study was inspired by Bieda's (2010) examination of the enactment of proof-related tasks in middle school classroom. Bieda (2010) used the mathematical task framework (Stein & Lane, 1996) to define a process of enactment of proof-related tasks in middle school classrooms as distinct from other types of "doing mathematics" tasks—i.e., tasks with high cognitive demand. A qualitative analysis of discourse around proof-related tasks in middle school classrooms found that the enactment of the proof-related tasks was generally not sufficient for students to learn to develop arguments that consistently satisfied Stylianides (2007) definition of proof.

Other studies that have described the activity of proving in classrooms paint in broad strokes the ritualistic aspects of proving that play out in the classroom—e.g., a teacher's narrow focus on the written form that an argument takes (Schoenfeld, 1988)—or parse the activity into coarse segments (e.g., the definition-theorem-proof sequence described by Weber (2004)) during which written board work is the primary mode through which the proof is developed and displayed. Studies that provide descriptions of the typical practices that students and teachers use when doing proofs in classrooms have been few. Furthermore, taking the institutional system of schooling into account when considering proof and proving has been the exception, rather than the rule, for studies of proof and proving in mathematics classrooms.

One study that does describe what actually happens in classrooms where proving takes place is a study by Herbst and Brach (2006) of doing proofs in high school geometry. Herbst and Brach (2006) recognize that “students experiences of doing math in school might be different from those of mathematicians who solve problems to create knowledge” (p. 75). That students’ experiences of mathematics in school would be different from those of mathematicians is linked to the institutional and instructional dimensions of mathematics as a school activity (Herbst & Brach, 2006, p. 74). Herbst and Brach (2006) do not evaluate this difference as an infelicity that needs to be cured, but rather recognize it as a basic fact of school mathematics that researchers can study. This perspective recognizes that there is an activity that students and teachers in geometry classrooms co-create that is recognizable (to students and teachers) as an occasion during which proofs are done. Herbst and Brach (2006) use interviews with students to uncover the qualities of this activity that make it recognizable to students in geometry classrooms. They suggest that there are norms of the doing proofs situation that teachers and students share and mutually uphold through fulfilling their roles in the situation. Investigating such norms of instructional situations can bring to light the routine features of instruction that would otherwise be transparent. They argue that identifying the routine features of mathematics instruction is a means of studying the practice of mathematics teaching.

The differences between the disciplinary and schooling dimensions of learning mathematics are pronounced in the case of proof. For mathematicians, "proofs are products of intellectual creation" that allow them to make important claims, and "any general notion of proof is just a name they use for their main methodological tool" (Herbst & Brach, 2006, p. 75). This perspective on proof echoes Rav's (1999) thesis that mathematicians create proofs to develop techniques to further the field of mathematical enquiry. But neither building new disciplinary knowledge nor furthering the methods of enquiry are at stake when students engage in proving in schools. Rather, for students, a proof is a means through which they may come to know some mathematical idea. But beyond this, a proof is "something that students are expected to learn in the high school geometry course; a general notion of proof is for them an end in itself" (p. 75). Sierpinska (2005) goes so far as to declare that,

> it is not realistic to expect that students will be 'initiated' into the discourse of working mathematicians; they can only be initiated into the discourse of school mathematics, where proofs are texts written in a distinct genre, having little to do with the truth of statements the students are supposed to prove and even less with communication of a result of an investigation. (Sierpinska, 2005, p. 9)

Sierpinska (2005) offered this assessment in an article that considers the turn toward discourse in mathematics education. The point of this critique is not to paint a hopeless picture for the prospect that mathematical novices (i.e., students) can be taught to develop into mathematical experts (i.e., mathematicians), but rather to temper the expectation that the generation of classroom discourse—any classroom discourse—is, itself, sufficient to bring about this initiation into mathematical culture. Elsewhere in the article, Sierpinska (2005) argues that mathematicians communicate in discipline-specific ways. What research can do is identify these discipline-specific communication practices and compare them to the communication practices that are developed in school mathematics classrooms. The current study contributes to this work by describing the communication practices that teachers recognize as routine when students are doing proofs in geometry.

That mathematicians generate proofs of their own (Rav, 1999) and check the proofs of others (Weber, 2008) has been well-covered in the literature. These activities have to do with the role that proof plays in building mathematical knowledge (generating proofs) or in honing one's own tools for creating proofs (proof validating, wherein mathematicians can study the techniques that make a proof work). Still a different activity that has garnered less attention in the literature, but that nonetheless might indicate practices around proofs that students might learn to develop, is the work of presenting or explaining a proof with which one is already familiar.

In the next section I consider accounts of the role that proof plays in explaining mathematical ideas. I argue that such accounts have framed proofs as entities that either do or do not carry explanations. Such a role for proofs as explainers overlooks the possibility that the act of carrying out an explanation—i.e., through a multimodal enactment of a proof—is an occasion for the person that is doing the explaining to hone proof-related mathematical practices. Presenting proofs in mathematics classrooms could create opportunities for novices to be initiated into disciplinary specific practices for communicating mathematics.

Disciplinary practices for communicating proofs

Apart from issues of determining what should count as a proof or whether students (or teachers) have a clear concept of mathematical proof that allows them to differentiate proof from mere empirical argument, there is a cluster of research on proof and proving that concerns the role that proof plays in explaining, understanding, and communicating mathematics. Hanna (2000) reflects that, after careful consideration, issues of the appropriate level of rigor of a proof are secondary to the role that proof plays in generating understanding. Hanna (2000) writes: "It became clear to me that a proof, valid as it might be in terms of formal derivation, actually becomes both convincing and legitimate to a mathematician only when it leads to real mathematical understanding" (p. 7). Hanna (1990) argues that, in school mathematics, the primary purpose of proof is to explain and that those proofs that provide explanations should be valued most highly (Hanna, 2000, p. 8). Hersh (1993) also argues that the principal role for proof in mathematics classrooms is to explain: a proof of a theorem should "provide insight into why the theorem is true" (p. 396). Proofs have educational value in so far as they are carriers of "complete understanding" of the mathematical ideas at stake in the proof (Hersh, 1993, p. 398). The teacher's role is to decide whether a particular mathematical idea calls for an explanation. If it does, the teacher can achieve the goal by offering an explanatory proof of the idea.

The explanatory potential of proofs is no doubt one reason for the prominent place the presentations of proofs play in university mathematics instruction (Weber, 2004; Artemeva & Fox, 2011). The presentation of mathematical proofs is a standard classroom activity that occurs in mathematics departments at universities throughout the world (Artemeva & Fox, 2011). The presentation of proofs in mathematics classrooms is a skill that mathematical experts develop as they hone their disciplinary communication practices (Artemva & Fox, 2011). Descriptions of how mathematical experts enact what has been described as *chalk talk* (Artemeva & Fox, 2011) are similar in different countries and different subjects (Greiffenhagen, 2014; Núñez, 2009).

Chalk talk is a form of mathematical communication during which a mathematical expert develops a mathematical work in real time for an audience (Artemeva & Fox, 2011). The mathematician engaged in chalk talk writes words, symbols, and draws pictures (or makes other inscriptions) while simultaneously talking and gesturing about what is being written (Artemeva & Fox, 2011; Greiffenhagen 2014; Núñez, 2009). The talk that accompanies the written development of a proof includes both verbalizing[6] the steps of the proof as they are being written and also providing commentary about the process/strategy of the proof as it unfolds. The coordination of speaking, writing, and gesturing, and even the sheer volume of writing that is produced during *chalk talk* are signature features of this mathematics-specific communication practice (Greiffenhagen, 2014).

Even in an age where it would be convenient to digitally project completed proofs during the course of instruction, university professors prefer *chalk talk* (Artemeva & Fox, 2011; Greiffenhagen, 2014) over projections. Whereas the projection of a proof would present the final, finished form an argument as a static whole, developing a proof via *chalk talk* provides an opportunity to demonstrate some—though certainly not all—of the reasoning practices that were used to generate the proof. The preference that mathematicians have for *chalk talk* is consistent with the descriptions of different cultures of proving that are provided by Livingston (1999), where the written records of proofs (i.e., a completed proof that could be projected) are descriptive traces of the spectrum of practices that the generator of the proof employed to establish the truth of the claim that is proved (Herbst & Brach, 2006, p. 73).

Mathematicians routinely engage in *chalk talk,* whether during the course of classroom instruction or when making presentations at professional meetings (Greiffenhagen & Sharrock, 2011; Weber, 2004). Greiffenhagen and Sharrock (2011) draw on Hersh's (1991) analysis of the 'front' and 'back' ends of mathematics as a discipline to analyze the relationship between proofs, mathematical knowledge, and the representation of mathematics as a field of inquiry. According to Greiffenhagen and Sharrock (2011), Hersh (1991) calls the polished, sequential, written mathematical arguments that are published in textbooks and journals the 'front' of mathematics.

[6] Artemeva & Fox (2011) do not home in on the differences between what mathematical experts say versus what they write when doing chalk talk, but there are differences.

The front of mathematics is what is presented to the world. Keeping up the appearance of pristine, certain knowledge is one of the primary ongoing achievements of the genre of professional mathematical writing. The 'back' of mathematics, on the other hand, is the actual day to day work of conjecturing, following false leads, making mistakes, using pictures, being stuck, waving one's hands to get unstuck—in short: the work that living breathing mathematicians conduct in their daily lives to get a problem to the point where it can be brought out to view in the front of mathematics.

Greiffenhagen and Sharrock (2011) complicate the clean distinction between the 'front' and 'back' ends of mathematics. They analyze the transcripts of university professors presenting proofs to their students in two different settings: at the blackboard (during a lecture) and a one-on-one meeting with a doctoral student. While it is the case that the written records of the proofs that end up on the blackboard during the lecture conform, generally, to the notion of the 'front' of mathematics that is advanced by Hersh (1991), the professor does more than simply write the proof on the board. In fact, the professor generates the proof by talking as he writes, and as he talks, the professor is explaining the key ideas that motivated the proof, the techniques that are being used to structure the argument, and he provides other convenient or timely details that help the argument that is being generated cohere (Greiffenhagen & Sharrock, 2011, pp. 10-13). As opposed to locating the agency for the explanation in the written record of the proof, Greiffenhagen and Sharrock (2011) consider the ways in which the mathematics professor's enactment of the proof—even one that could be in the showroom of the 'front' of mathematics—is an explanation. Presenting the proof to the class during the lecture calls on the agent doing the presenting—in this case, the professor—to generate an explanation of the proof as it is being written. This regeneration of a proof before an audience is the essence of *chalk talk.*

When examining what happens in the 'back' of mathematics during the one-on-one meeting with the doctoral students, Greiffenhagen and Sharrock (2011) found consistency between the kinds of explanatory moves the mathematician made to motivate the proof that was presented during the lecture and his efforts to understand the still-in-progress proof that was being developed by the doctoral student. For example, in both cases, the professor verbalized while working through the steps that he had written down. The written steps carried the necessary formal machinery of the proof, while the professor's speech focused on the strategic and conceptual flow of the argument in which he was engaging. Greiffenhagen and Sharrock (2011) argue that Hersh (1991) used the 'front/back' divide to suggest that the front of mathematics is a ruse, one that allows the myth of mathematical certainty to be preserved amidst the non-technical public. Their analysis of the practices that unfold in the display window (lecture) and back rooms (one-on-one meeting) of mathematics suggest, however, that in each setting, there are material, non-technical practices (gestures, drawings, comments to oneself) that mathematicians use when explaining proofs. That is, even while proofs are still being developed, mathematicians engage in communication practices of *chalk talk* to understand where a proof currently is and where it might still need to go.

Presenting and explaining proofs in schools

Because of their role in establishing and archiving mathematical knowledge, "doing proofs" has come to be almost synonymous with "doing mathematics". But *doing proofs* (in school) and *proving* (in mathematics) are different kinds of activities (Herbst, 2002; 2006), and the differences between them seem to be ineluctable (Sierpinska, 2005). Doing proofs in school is a didactic game in which certain kinds of specified tasks may be traded for particular knowledge claims (Herbst, 2006). By contrast, the comparable activity in the discipline is a knowledge-generating enterprise whose goods are written records of proofs (Livingston, 1999) that have primarily archival value (Greiffenhagen & Sharrock, 2011).

Mathematics education research about proof and proving has focused on epistemological and psychological aspects of proof. Epistemologically, the challenge has been to define proof in schools in a way that is functional for classrooms but still consistent with disciplinary exemplars of proof. Psychologically, the challenge has been to identify the conceptual schemes that individual students use to process the idea of mathematical proof and to try to define instructional strategies that can bring those schemes more in line with the proof schemes of mathematicians. When the role of proof in communication and explanation is considered in the literature, what is at stake is: (1) the extent to which the proof, as an entity, is a suitable transmitter or container for a mathematical explanation (the proof that explains), or (2) the status of the proof (valid/invalid, counts/does not count) as something that is socially negotiated among a community of mathematicians—i.e., what "counts" as proof depends on what will convince a community of mathematical experts.

What is missing from these perspectives is a consideration of the possibility that the capacity to explain a proof in a public setting—where one has the opportunity to coordinate speech, writing, drawing, and gesturing during the enactment of the proof—is an occasion to develop discipline-specific communication practices. Investigating the practices used for explaining proofs could be generative for the study of proof in mathematics education because the practices mathematicians use to explain familiar proofs are similar to the practices they use to generate novel proofs (Greiffenhagen & Sharrock, 2011). The theory of instructional exchanges presented in the next section provides a framework for investigating the classroom practices through which proofs are presented and explained.

Theoretical Framework

A motivating example

Mathematics classrooms are institutional settings in which teachers and students interact around mathematics content (Cohen, Raudenbush, & Ball, 2003; Herbst & Chazan, 2012; Kilpatrick, 2001). The institution of school is a mediator between the universe of knowledge claims available in a field (e.g., mathematics) and the students in a class. An all-too-common classroom event illustrates the institution's mediating effect.

At some point in a career, any ambitious teacher will be asked by a student: "Is this going to be on the test?" For reasons that will become clear, I call this the *bookkeeping question*. Teachers tend to be asked the bookkeeping question when what they are teaching veers into the non-routine, such as during those instances when a teacher momentarily transcends the institution and offers students a glimpse of what a field of study is really like. But even during such moments, the students and teachers are still in school. And the bookkeeping question is a way for the students to re-establish the institutional terms of the activity in which they and the teacher are engaged.

Students ask the bookkeeping question when it is not clear whether they are in the realm of knowledge they will be held accountable to produce in the future or whether what the teacher is showing them is just "something extra" that the teacher happens to find interesting. Certainly, in any class there are students who appreciate the opportunity to learn something because of the intrinsic value of finding things out. But the reality that some students are genuinely interested in the knowledge for its own sake does not alter the institutional circumstances of the learning that takes place in school: There are topics that are in the curriculum that students are required—by the school, the state, or a national school governing body—to learn and that teachers are required to teach. The *bookkeeping question* thus places the teacher in a double-bind, not unlike the one described by Herbst (2002) when engaging students in proving. On the one hand, if the teacher says “no”, students invariably take this as a license to all but put their heads down and go to sleep[7]. By telling the students that “what I am now showing is *not* something I am going to ask you about later”, the knowledge that the teacher shares no longer has any institutional standing in the instructional activity in which they are engaged. On the other hand, if the teacher says “yes”, the teacher is no longer giving the students a glimpse of the field for its own sake—which was the teacher’s point in going off on the tangent in the first place—but rather is adding to the burden of institutionalized knowledge that the students are compelled (because they are students in a school system) to acquire.

The bookkeeping question betrays students’ awareness of their roles in the mathematics classroom: Students are asked by the teacher to produce work in exchange for recognition (from the teacher) that they have learned what they are required to learn (Herbst & Chazan, 2012). Students ask the bookkeeping question when they suspect that they are no longer in a familiar situation (Herbst, 2006), and they are attempting to clarify the terms of the activity in which the teacher has engaged them. In this way, the bookkeeping question functions as a corrective measure. When applied by students, it effectively reminds a teacher—i.e., brings the teacher back down to earth—that they are not just gathered in this place of their own free will to learn what the teacher has to teach (Herbst & Chazan, 2012, p. 605), but that they are enacting the role

[7] This is an exaggeration. Without hyperbole, what I mean to say is that, a central means that students in schools demonstrate that they have learned what they are required to know is through how students fare on tests (Nolan, 2012). If a teacher says: “what I am about to tell you is not going to be on the test”, the teacher is effectively saying that the students are not going to be accountable to demonstrate (on a future test) that they have learned what the teacher is about to teach. I am not suggesting that all students would immediately lose interest in what a teacher is saying simply because they will not be tested on it. What I am saying is that the *bookkeeping* question makes the transactional nature of school instruction explicit.

of student and are institutionally accountable for knowing particular things. The bookkeeping question underscores the transactional nature of instruction. The question asks: "Will we have to produce work for that?"

The tension encapsulated above between the work students do and the knowledge claims that are available underscores the transactional nature of the activity that takes place in school mathematics classrooms. Sierpinska's (2005) observation (quoted above) that school mathematics initiates students into the practices of school (rather than the practices of the discipline) resonates with this account. Strategies for helping students develop disciplinary communication practices can use the transactional nature of school mathematics as a design principle.

Research that focuses on getting students to recognize what a proof is in the discipline rests on an expectation that students are willing and able to see through the didactic game they are being asked to play whenever they are required to do proofs in schools. The nature of the game becomes especially evident in situations where students are being asked to prove a result they already know, usually with the teacher making some effort to emphasize that the results they already know are not admissible in the proof. Rather than deny that the game exists, an alternative strategy is to describe how the game is played. An advantage of describing instructional transactions as they actually occur is that such descriptions can be a basis for designing interventions to improve instruction that contend with the social realities of classroom activity. In the case of mathematical communication, describing the communication practices students actually use when working on tasks that are typically assigned is a way to shed light on opportunities for engineering tasks that would call for different kinds of communication practices. A model of classroom activity that provides a framework to situate such descriptions is the theory of instructional exchanges (Herbst, 2006; Herbst & Chazan, 2012).

Instructional exchanges and instructional situations

Instructional exchanges (Aaron & Herbst, 2012; Herbst, 2006; Herbst & Chazan, 2011) are acts through which students trade work for claims—from the teacher—that they have acquired particular items of knowledge. An example of an exchange that is seminal to work conducted for this dissertation is the exchange of work on proving a specific proposition for the claim that students know how to do proofs. One of the items of knowledge at stake when students are learning to do proofs in geometry is learning to write an angle addition argument. In the language of instructional exchanges, a teacher achieves the teaching of this knowledge by assigning to students a proving task that requires students to use this technique to produce a proof. Students stake their claims on this knowledge by producing a proof that the teacher accepts. Although the work that students do on the task is particular—i.e., the proving task the teacher assigns to the student involves a specific claim—the knowledge at stake in the transaction is more general.

The activity that unfolds in mathematics classrooms can be described as a set of instructional exchanges. This way of conceptualizing the activity of mathematics classrooms draws on the notion of a symbolic economy (Herbst, 2006 citing Bourdieu, 1998). The implicit rules of this symbolic economy—in which work is exchanged for knowledge claims—comprise *a didactical contract* (Brousseau, 1997; Herbst, 2002; Herbst & Chazan, 2011) that students and teachers tacitly accept and mutually uphold. The didactical contract "establishes global responsibilities between teacher, student, and subject of study" (Herbst, 2006, p. 315). A fundamental element of this contract can be expressed through the rule that "the teacher helps the student study the content" (Herbst & Chazan, 2012, p. 600).

The rule that the teacher helps the student learn the content helps to structure the exchanges of work for knowledge that take place in classrooms. In enacting this rule, a person playing the role of the teacher regulates the exchange by (among other things) deciding what knowledge is at stake in a particular exchange, what kinds of student work may be exchanged for that knowledge, and whether any specific instance of work satisfies the terms of the exchange. Students also have roles to play in an exchange (Aaron & Herbst, 2012): a person playing the role of student participates in the exchange by (among other things) producing the required (i.e., expected) work in a manner that the teacher is likely to accept.

The theory of instructional exchanges provides a global structure that describes how work produced by students is exchanged or *cashed* (Herbst & Chazan, 2012) for knowledge claims. Complementing this global structure, particular instructional exchanges have a local structure that makes them recognizable as instances of an *instructional situation* (Herbst, 2006). Instructional situations are identifiable by the routine activities of teachers and students within particular exchanges that belong in that situation. These distinguishing activities are *routine* in the ethnomethodological sense (Garfinkel, 1963) of being stable, reproducible, and generally transparent unless they are breached. Mathematics education researchers (Herbst, 2006; Herbst & Chazan, 2012) have described the routine activities that occur in instructional situations as *situational norms*.

Instructional situations are recognizable by the knowledge claims at stake within the situation and the type of student work the teacher will accept in exchange for those claims. The global structure of instructional exchanges and the local structure of situational norms provide heuristics for recognizing—amidst the totality of activity that takes place in a geometry classroom—the activity of *doing proofs in geometry*. I say "heuristics for recognizing" rather than “characterize” or “define” to highlight the fact that I am not proposing a set of rules for defining the instructional situation of “doing proofs”. Rather, I am making an argument that even without such an explicit set of rules, there are aspects of the activity of doing proofs—i.e., the norms of the situation—that are sufficient to identify when something is or is not an instance of this situation.

Instructional situations are stable segments of classroom activity in which teachers and students comply with expected ways of fulfilling their roles in the exchange that takes place within the situation. It is also possible that teachers or students will depart from what is expected while an exchange is taking place. The bookkeeping question described above is a student response to a generic example of such a departure. Specific examples of departures from what is normative could include a teacher posing a problem that has no solution, or a teacher cashing work on grounds that appear arbitrary, such as the quality of a student’s handwriting[8].

An instructional situation is “a system of norms adapted to regulate exchange of work on particular kinds of tasks for particular objects of knowledge” (Herbst & Chazan, 2012, p. 600). The kinds of tasks that a teacher assigns for students to do, the kinds of knowledge available, and the work that students are to produce in response to the task are carried out in accordance with expectations that are shared by teachers and students. These expectations are *situational norms*—i.e., situation-specific regularities of the activity that takes place in mathematics classrooms. In the instructional situation of doing proofs in geometry, researchers (Herbst et al., 2009; Herbst, Aaron, Dimmel, & Erickson, 2013) have identified situational norms that concern how teachers specify the proof task that is to be completed, the division of labor between teacher and students

[8] There are situations where penmanship is the knowledge at stake, such as when learning to write in elementary school. In other contexts, though, a teacher focusing only on the quality of penmanship would be a breach of the tacitly agreed upon terms of the exchange.

with respect to generating the proof (e.g., the teacher provides—and, if necessary, modifies—the diagram, the students provide statements and reasons in sequence), and norms for the mathematical symbol system in which the proving work takes place. The research conducted for this study examines how teachers expect students to represent their work on proofs in order for those proofs to be exchanged for credit. This calls for describing the semiotic terms of the exchange.

Semiotic terms of the exchange of proof for knowledge claims in geometry

When doing proofs in geometry, the activity of exchanging a proof for knowledge claims involves checking a proof to determine whether or not the proof is acceptable. The question for the teacher—about whether or not a particular instance of proof should be cashed—is a different question than the question for the researcher about whether or not a student has generated an argument that satisfies a prescriptive definition of proof. Although a teacher is looking for evidence that a student's proving work can be cashed, what exactly the teacher is looking for—and what the teacher finds—are descriptive questions. That is, doing proofs in geometry happens regardless of teachers' conceptions of what proofs actually are, and it happens regardless of whether students discern a clear and mathematically coherent concept of proof (from the perspective of a researcher) as a result of the exchange. The eventual goal of doing proofs in geometry might be for students to develop such an understanding of proof, but the attainment of this goal is mediated through work that students can actually do.

A semiotic description of the terms of the exchange of proof for knowledge claims when doing proofs in geometry will necessarily involve a range of communication modes, because

> The work that students do is multimodal: it involves strokes on paper, utterances, silences, gestures, physical movements, etc. Many of those moves do not have conventional names, let alone mathematical names—some of them may not even be noticed if done outside the sequence of actions where they ordinarily appear. On the other hand the knowledge at stake in an instructional exchange (e.g., solving equations in one

> variable) is, in general, stated through a combination of elements of the mathematical register (using language, symbols, and sometimes diagrams) all of which have some degree of stability and status in adult mathematical discourse. All knowledge at stake can be stated explicitly in some way, yet its transaction, in exchange for work done, requires the teacher to engage in serious interpretation of much more opaque signs in the realm of the students' work (Herbst & Chazan, 2012, p. 601)

Herbst and Chazan (2012) describe in a general way the multimodal production of semiotic resources by students when working on a task to fulfill the terms of a generic exchange. But the observation, in general, that student work can be multimodal does not mean that, in all exchanges, teachers have developed routines for noting and giving value to the multimodal work that students produce.

An example where multimodality plays no role in the exchange—though it may play a role in the work—is a written exam. It certainly is the case that teachers have strategies for giving value to a range of formal to informal work that is written on exams, including drawings, pictures, and even pleas for mercy[9]. But it is also the case that, on an exam, how a student gazes at the page when reading a question, or the order in which a student answers the questions on the exam—each of which could provide information about the work the student is producing[10]—are not resources that can be accounted for in the exchange, because written examinations facilitate exclusively written exchanges.

Historically, exchanges of proofs in geometry have been facilitated through the use of the two-column format (Herbst, 2002). A two-column proof is a structure for organizing the statements and reasons that link information that is given to a claim to be proved. The referents for the statements in the proof are objects that are represented in a diagram that accompanies the

[9] As in: "I know that I need to use angle addition but I can't figure out what to add"
[10] As an example: There are occasions where strategic test takers can use the information on a later question to return to or revise their answer to an earlier question. There may be a trace of this in the fact that a student erased or otherwise amended the earlier answer, but the temporality and causality of the modification is not captured by the written record of the exam.

proof. For proofs that are exchanged in writing—such as on a homework assignment, or on an exam—the analysis of the semiotic terms of the exchange involves describing the details that teachers are looking for in a proof. This will be considered in Chapter 3 when teacher routines for checking proofs are described.

Proofs can also be exchanged during the activities that take place in geometry classrooms. For example, a student could be called to the board to present a proof to the class. In such a circumstance, knowledge of how to explain a proof through a presentation could be part of the knowledge that is at stake in the exchange. A student presentation of a proof has the potential to involve more than just the written statements and reasons of the proof, but could also involve the words that a student says, the movements a student makes, and the additional marks that a student could add to the diagram (e.g., congruence markings) or the argument (e.g., underlining a key step). Describing how opportunities for students to use these other modes of mathematical expression play out in mathematics classrooms is the purpose of the study reported in Chapter 3.

Chapter Summary

Research on proof and proving in schools has been preoccupied with determining what should or should not count as proof in schools from an outsider's perspective (e.g., mathematics education researchers, mathematician). The focus in the research on proofs as epistemological entities that have a special status underpins much of the research on students' and teachers' conceptions of proof and research that is concerned with an appropriate definition for proof in schools. The entity focus on proof has deepened our understanding of the individual obstacles students and teachers face to developing a concept-image of proof that aligns with that of mathematicians. However, the proof-as-entity focus has also diverted attention from describing how proof and proving unfold as typical classroom activities.

The research conducted for this dissertation aims to describe how modes of communication are used during the classroom activity of doing proofs. In providing such a description, I draw on previous descriptions of the doing proofs situation (Herbst et al., 2009) as a norm-regulated social activity. The notion that there are situational norms when doing proofs in geometry suggests that, within the situation, there could be normative ways that mathematics is communicated. The research conducted for this study aims to uncover such norms.

Preview of the studies reported in Chapter 3 and Chapter 4

Chapter 3 provides the foundation for the empirical study that is described in Chapter 4. In Chapter 3, I analyze video episodes of students presenting proofs and teachers checking proofs (i.e., determining whether or not a proof will be cashed by the teacher in the situation) to provide descriptions of the semiotic resources that teachers and students use when doing proofs in geometry. I find that teachers and students use semiotic resources in different ways, and that the only semiotic resources that have currency in the situation—from the standpoint of representing student work that can be accepted by the teacher for knowledge claims—are the written statements and reasons of a proof. That writing plays such an important role in representations of student work on proofs is neither surprising nor unreasonable, given the historical and disciplinary connections between mathematics and the making of inscriptions (Davis, 1972; O'Halloran, 2005; Greiffenhagen, 2014). My goal in describing the expectations that teachers have of student mathematical work is not to judge those expectations but simply to identify them, to bring them to the surface as features of classroom activity that happen.

The semiotic norms I hypothesize at the end of Chapter 3 are rooted in an observer's perspective on the regularity of the activity that unfolds in the situation of doing proofs in geometry. Determining whether such norms are recognizable from the perspective of a secondary teacher is the aim of the empirical study described in Chapter 4. The design of the study is a variant of the virtual breaching experiment (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009; Herbst, Aaron, Dimmel, & Erickson, 2013) technique that has been used to investigate whether teachers recognize hypothesized situational norms. The purpose of designing and conducting the experiment was to determine the extent to which the semiotic norms that I observed in the episodes of doing proofs in geometry reported in Chapter 3 are recognized by secondary teachers. Measuring the degree to which teachers recognize a hypothesized norm is significant because any changes to classroom instruction must first contend with the normative ways the instruction unfolds in classrooms (Herbst, 2006). By identifying what teachers recognize to be normative—and how teachers react to representations of classrooms in which the norm is breached—indicates what it might cost to change the norm.

CHAPTER 3:

A MULTIMODAL DESCRIPTION OF DOING PROOFS IN GEOMETRY

In this chapter, I provide descriptions of student work in instances of the situation of doing proofs in geometry. The situation of doing proofs could be realized through a range of activity types (Lemke, 1990). For example, students might work on proofs individually during class, generate a proof collectively through a teacher-led discussion, or work on a proof problem in small groups. *Working on seat work, participating in whole-class discussion,* and *working in groups* are different types of activities in which students engage when fulfilling their roles as students. The situation of doing proofs could combine with each of these activities and such combinations would yield different realizations of the situation. In this way, instructional situations are parallel to *genres* of social activity (Martin & Rose, 2003; Kress, 2003).

A genre is a staged, goal-oriented social activity. Viewing instructional situations as genres provides a conceptual footing for describing how the *modes* of communication (Kress, 2009) available in the situation are used by teachers and students. The term *mode* is used with various meanings in the literature on language and communication. In what follows, I use *mode* to refer to the *materials* (e.g., inscriptions, sounds, gestures) of representation that are selected, used, and shaped (over time) by a culture (Kress, 2009). Below, I identify *modes* that have been selected and shaped by cultures of mathematical experts.

Overview of multimodal descriptions of doing proofs

I present descriptions of video episodes of students *presenting* proofs and teachers *checking* proofs below. The goal is to describe how teachers expect students to use different modes of communication. The descriptions are multimodal in the sense that I attempt to capture how a range of communication modes are used in the episodes that are described. The *presenting proofs* realization of doing proofs is one in which a proof is presented to the class. The analysis of episodes of presenting proofs reported below casts students in the role of doing the presenting. The *checking proofs* realization of doing proofs is one in which a completed proof is scrutinized to verify that it is complete and acceptable. The analysis of episodes of checking proofs reported below casts teachers in the role of leading the proof checking.

Analyzing instances of the *presenting* and *checking* proofs realizations of doing proofs provides an occasion for zooming into how teachers manage the semiotics of the exchange of work for knowledge claims when doing proofs in geometry—e.g., How do teachers and students differ in their use of semiotic resources? What expectations do teachers have for how students will present a proof to the class? What do teachers require of a proof for it to be acceptable? Such an account of the doing proofs situation complements prior research that has described how proof tasks are posed (Herbst et al., 2009; Herbst, Aaron, Dimmel, & Erickson, 2013) and how proofs are generated (Herbst et al. 2009) when doing proofs in geometry.

To frame the descriptions, I identify different modes of communication that are available when doing proofs in geometry. The modes I identify and their systemic organization is theoretically grounded in the social semiotic literature on multimodality. I establish links to this literature as I describe the modes for communicating when doing proofs.

Following the consideration of the different possible modes, I report the descriptions of the video episodes. I start by reporting the descriptions of students presenting proofs to the class. After describing these episodes, I move to the the descriptions of teachers checking proofs in geometry class. The activities of *presenting* and *checking* proofs were purposefully selected from a corpus of video records of geometry classrooms (described below). I focused the descriptions on these realizations of doing proofs because they represent opportunities for students in geometry classrooms to engage in discipline-specific communication practices.

Rationale for investigating the *presenting* and *checking* realizations of doing proofs.

Presenting proofs to the class could provide an opportunity for students to develop the discipline-specific communication practice of *chalk talk.* Identifying the communication practices that teachers expect students to use when presenting proofs to a class provides a means for describing how teachers help students develop their multimodal mathematical communication skills. When students are presenting proofs to a class, the teacher has greater access to the student's non-written modes of communication. This is not necessarily the case in other realizations of the doing proofs activity.

For instance, when students work on proofs in groups together during class, non-written modes of communication no doubt play a part in the work that students do to generate the proof. But during this activity, there is almost no practical way for a teacher to asses how students in different groups across the classroom are using these non-written resources in real-time. When a student is called to the board to present a proof, or to present and explain a proof, however, it is more realistic that non-written modes of communication could play a part in what teachers look for in an exchange.

Checking proofs could provide an opportunity for students to engage in the discipline specific practice of determining whether a proof is valid (Davis, 1972; Weber, 2008). When teachers lead the class through checking a proof, the nature of the activity is one where the students and teacher are negotiating together whether the proof that has been produced is acceptable. The goal of checking a proof is to make the terms of the exchange—from the teacher's perspective—visible to the students. I describe episodes of proofs being checked to uncover the customary expectations teachers have for how students use written and diagrammatic resources when doing proofs.

I conclude this chapter by distilling some of the regularities that I observed across the video episodes of presenting and checking proofs into *semiotic norms*. A semiotic norm is a particular kind of situational norm that describes how semiotic resources are expected (by teachers) to be used within the situation. The semiotic norms that I identify are conjectures about the routine ways semiotic resources are used when proofs are presented and checked in geometry classrooms. The conjectures are warranted by the descriptions of the video episodes. The purpose of the study that is described in Chapter 4 is to determine whether these conjectured routines are recognizable to geometry teachers.

Modes for communication when doing proofs in geometry

One entry point for considering the communication modes that are used when doing proofs is to begin at a general level with a consideration of the physical aspects of the space where the activity of doing proofs occurs. Mathematics classrooms provide a physical setting for the kinds of semiotic resources that can be produced when doing proofs in geometry. Though classrooms may differ in their particular spatial arrangements—e.g., there are hard differences in size, shape, and other infrastructural qualities, and soft differences such as the configuration of desks and other furniture, or what is displayed on the walls—a nearly universal feature of the physical spaces in which the activity of doing proofs occurs is its organization around a central space for creating and displaying inscriptions, such as a chalkboard or whiteboard. Spaces for making inscriptions are indispensable in mathematical cultures, be they the sand flats used by Archimedes (Davis, 1972) or the generous, multiple blackboards that are part of the institutional identity of university mathematics classrooms (Greiffenhagen, 2014). I will refer to a generic space for displaying inscriptions as a *canvas* and the canvas-appropriate tool with which inscriptions are created as a *stylus*.

An initial way to organize the communication modes that agents—enhanced with a stylus and canvas—can produce when doing proofs is to catalogue them according to the sensory channels[11] (Hirsh & Sherrick, 1961) through which they are created and apprehended. Figure 1 shows a systemic organization of some of the semiotic resources of *doing proofs in geometry*.

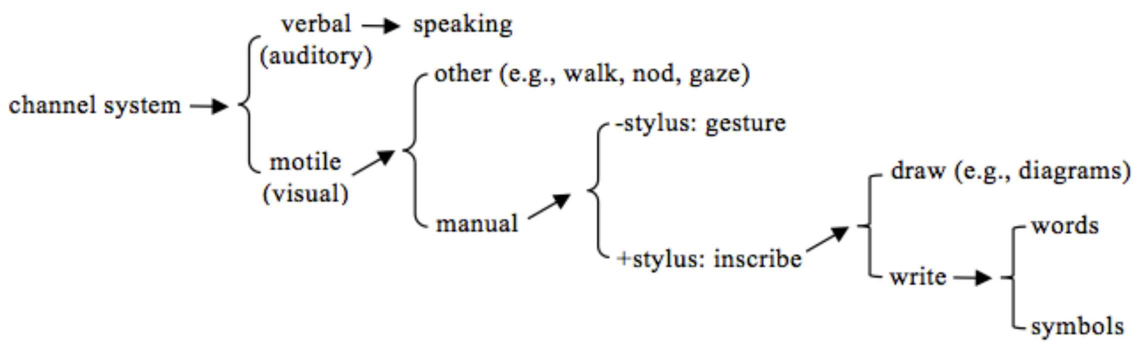

Figure 1: Semiotic network representing the communication modes available when *doing proofs in geometry*.

Figure 1 represents the modes of communication that are available when presenting a proof, systemically organized by sensory channel and developed specifically for the motile (visual) subsystem. The network is systemic in the sense that it spans the set of possibilities for semiotic resources in a doing proofs situation. These possibilities are arranged as choices within the network—curly brackets indicate inclusive-or choices and square brackets to indicate choices that are exclusive-or.

[11] I use *sensory channel* in a deliberate departure from *sensory modality* (a standard term for referring to the means through which our senses process input) in an effort to maintain a clear distinction between the capacity for producing/apprehending semiotic resources (traditionally, sensory modalities) and the materiality of those resources themselves (what I have called communication modes, following Kress (2009)).

The first choice is an inclusive choice between *verbal* and *motile*. This choice is inclusive because as an agent is creating semiotic resources by speaking, this agent could also be creating semiotic resources through moving, e.g., by gesturing while speaking. The inclusive choices continue through the different branches of the network, up to the choice between *write* (e.g., letters, symbols) and *draw* (e.g., diagrams), since only one of these resources may be produced at a time.[12]

As mentioned above, stylus and canvas refer to physical aspects of the space of a geometry classroom that facilitate the production and display of signs. A stylus is a means for making inscriptions, such as a piece of chalk or marker, while a canvas is a surface on which signs may be inscribed, such as a black or white board. Stylus and canvas are intended to be general enough to include technologies like dynamic geometry software—where the canvas would be the display screen and the stylus would be the keyboard, mouse, or other means through which an agent can create signs in the software—as well as classic tools like compasses and straightedges—each of which can be conceptualized as a type of restricted stylus: A straightedge is a stylus that can make arbitrary straight strokes, while a compass is a stylus that can make arbitrary circles. The different types of styluses make different physical (in the sense of wielding the stylus) and logical (in the sense of understanding what types of signs it can produce and how it produces them) demands on a person (or persons) using the stylus, and these differences may affect the use of resources from other modes.

Coordinating resources from these different modes (i.e., knowing what kind of diagram to draw when, how to speak and write or speak and draw simultaneously) is a non-trivial mark of one's multimodal literacy within a social activity (Kress, 2009). The coordination of speaking

[12] This assumes that a person does not normally write two different things simultaneously.

with writing, drawing, and gesturing is a feature of *chalk talk* (Artemeva & Fox, 2011), an activity in which a mathematical expert uses the range of modes shown in Figure 1 to develop a proof at a chalkboard. Describing the range of available semiotic resources in terms of how they are created or apprehended highlights the importance of accounting for differences in materiality among semiotic resources. According to Kress (2009), *modes* are the material resources involved in making meaning (p. 105). Accounting for the materiality of the various semiotic resources is an essential characteristic of multimodal descriptions of social situations:

> Mainstream linguistic theories of the twentieth century had emphasized abstraction and generalization…. A multimodal social-semiotic approach to *representation* by contrast puts the emphasis on the *material,* the *physical,* the *sensory,* the *bodily,* ‘the stuffness of stuff’, away from abstractions, toward the specific, the variable. (Kress, 2009, p. 105)

Figure 1 is organized by sensory channel. *Verbal* and *Motile* refer to channels that primarily produce semiotic resources, whereas *auditory* and *visual* refer to channels that primarily apprehend resources. Verbal resources include spoken words. Motile resources include written inscriptions (e.g., words, symbols) drawings, and gestures—specifically, movements of a speaker’s hands that are co-expressed with speech (McNeill, 2008). Gestures are semiotic resources that convey information that is non-redundant with the speech (McNeill, 2008).

McNeill (2008) identifies four dimensions along which the physical form of gestures can vary—these are the iconic, deictic, metaphoric, and beat categories of gesture. In addition to these dimensions of gestures, McNeill (2008) identifies *emblems* as a category of hand signals that are fixed in form and have meanings that are culturally stable. An example is the emblem for “OK” used in America (the tip of the index finger is tucked under the tip of the thumb, the remaining three fingers are extended).

Iconic gestures are icons of what they represent, such as making a triangle with one's fingers while one utters the word “triangle”. A deictic gesture is a pointing gesture, using one's hands, fingers, or other body part (which can be culturally dependent), often seen in the context of geometry when speakers are referring to “this line” (points at a stroke) as opposed to “that one” (points at another stroke). Metaphoric gestures are like iconic gestures, in that they are meant to represent some entity, yet they are different, in that how they represent the entity is not literal—people often gesture with their hands open when they are talking about “an idea”, as if whatever the idea were being held in the person’s hand (see Figure 2).

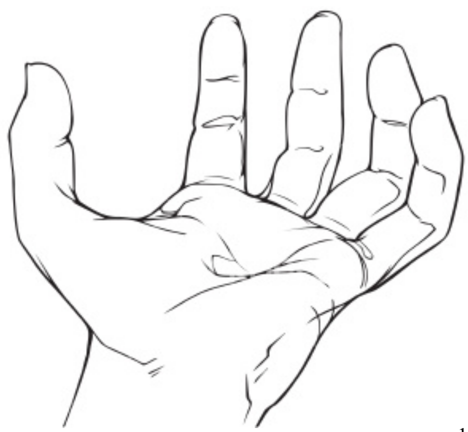

Figure 2: The “hand-open, palm-up” gesture[13].

This occurs regardless of the materiality (or lack their of) of the idea in question. Hence, a thunderstorm (meteorological idea) might be accompanied by a metaphoric gesture that looks similar to a metaphorical gesture that accompanies a different idea, such as that of “unbounded growth”. In either case, one could expect a speaker to use a “hand-open, palm up” gesture, as if holding some big, not-well-defined thing (Kendon, 1994). Finally, “beats” are rhythmic gestures that speakers use for emphasis, characteristically evident in speeches made by politicians—such as the “Clinton thumb”: Bill Clinton’s signature “vertical fist/thumb up” gesture[14] (see Figure 3).

[13] Image source: www.istockphoto.com
[14] Though this gesture bears a resemblance to a “thumbs up,” Clinton’s gesture is used rhythmically while he speaks and does not, by itself, convey any standard message. This image was the top hit on a Google search for “Bill Clinton thumb gesture.”

Figure 3: Bill Clinton using his signature beat.[15]

Descriptions of episodes of *presenting* and *checking* proofs

Teachers and students play different roles and have different levels of agency within an instructional situation (Aaron & Herbst, 2012). The description of instances of doing proofs presented below examines the ways that teachers regulate how students use different modes of communication when presenting proofs on a central canvas in class. The semiotic analysis reported below uses a combination of still images and multimodal transcription (van Leeuwen, 2005; Kress, 2009; Jewitt, 2009) to describe how semiotic resources are used by actors in the doing proofs situation. The principal systems that are considered below are those that are produced verbally (spoken words) and manually (both with and without a stylus).

The lesson corpus

The video episodes of geometry classrooms that are analyzed below belong in an archive of geometry lessons collected for an NSF[16] project that was active from 2002-2009. A subset of that corpus was analyzed for this study with the PI's permission. The lesson corpus contains

[15] Image retrieved from: sites.google.com/a/students.colgate.edu/gesture-brain-and-language/home/types-of-gestures

[16] The videos had been collected with resources from NSF grant REC 0133619 to Patricio Herbst. Neither the videotaped actions nor this commentary necessarily represents the views of the Foundation.

video records of four teachers (Cecilia, Megan, Lucille, and Emma[17]) teaching different sections of high school geometry. These were teachers working in Midwestern High School, a large public school in a Midwestern state.

The lessons were partitioned into episodes that were defined by activity structures (Lemke, 1990)—going over homework, teacher at the board explaining a proof, students working in groups—that occurred in the classrooms[18]. The original goal of capturing the classroom video was to provide records of routine, typical, or usual geometry classrooms. Prior analyses of episodes of the video corpus sought to surface the routine ways that teachers and students play their roles in the instructional situations that occur in geometry classrooms (see Herbst et al., 2009). These routines are examples of situational norms that were described above.

I surveyed the video corpus for instances of students or teachers doing proofs. Through this survey, I identified realizations of the doing proofs situation during which a teacher or student does a proof in front of the class—that is, enacts one of the stages of the genre of doing proofs identified above. This work to identify enactments of doing proofs proceeded iteratively, at increasingly fine-grained levels of analysis.

The first level of the review was a coarse screening of an episode to determine if there were any instances where students or teachers were engaged in any of the stages of the genre of doing proofs that were identified above. The first level of the review identified 54 candidate episodes. The second level of the review sorted the records from the first level according to the primary agent (student or teacher) that was doing the enacting. The third level of the review sorted the records from the second level according to the viability of analyzing the records for

[17] The video records of this teacher's class did not contain any instances of doing proofs and were not analyzed for this study.
[18] The organization of the corpus into episodes was conducted by previous members of the GRIP research lab.

evidence of multimodal communication—i.e., does the camera capture a sufficient portion of the activity so that the coordination of semiotic resources from different modes could be documented and described? After the final level of the review, there were 21 episodes. From these 21 episodes, I selected the segments of the episodes that are analyzed below. In selecting segments to analyze, I looked to capture a range of how the *presenting* and *checking* realizations of doing proofs unfolded in geometry classrooms. Thus, the selected segments constitute a *purposeful sample* of the episodes in the lesson corpus (Stylianides, 2007, citing Patton, 1990).

The episodes of students presenting proofs that are described below document the resources that students use when presenting proofs. The video episodes analyzed below provide evidence that when students present proofs in geometry, they are primarily accountable to the written record of the proof they produce on the board. A routine that illustrates this is one that I call *proof transcription,* during which a student creates a mark-for-mark reproduction of a proof on a whiteboard or chalkboard.

Episodes of students presenting proofs set up a contrast between the semiotic resources that are, in theory, available to students when presenting proofs and those resources that, in practice, they are actually held accountable for using. The third section of the descriptions of episodes of proofs being presented to geometry classrooms examines cases of teachers checking students' proofs to determine if a proof is acceptable. The purpose of the third section is to provide evidence that geometry teachers regulate the exchange of proofs for knowledge claims in normative ways that are linked to the mode of communication.

Method of description

The analysis of instances of doing proofs reported below aims to describe ways that different agents use the available semiotic resources when engaged in the activity of doing proofs. The descriptions of the episodes are multimodal in the sense that semiotic resources from verbal and motile channels have been documented (through screenshots) and described (through summary comments). Following the assumption that multimodal transcriptions can help understand the relationship between a specific instance of a genre and the genre's typical features (Baldry & Thibault, 2003, quoted in Jewitt, 2009, p. 45), the transcriptions of excerpts from proof performances by students reported below aim to semiotically characterize the activity that occurs during realizations of the genre of doing proofs. Specifically, my goal in providing these descriptions of how semiotic resources are used by students in geometry classrooms is to surface the expected or regular ways that students communicate about proofs in geometry.

A central issue in multimodal transcription is how to represent the different modes of communication—which are generally happening simultaneously—so that they may be analyzed without defaulting to historical biases that favor the linear, sequential order of written text over the spatial, holistic order of visual and other non-verbal modes of communication (e.g., visually reading a diagram) (Jewitt, 2009). According to the Routledge Handbook of Multimodal Analysis (Jewitt, 2009), this

> critical issue—namely, the representation of the complex simultaneity of different modes and their different structure and materiality—has not been resolved in transcription, nor have satisfactory ways as yet been found to combine the spatial, the visual, and the temporal within one system in ways that take account of the perceptual difficulties that inevitably arise when attempting a simultaneous reading (p. 46).

The descriptions of the instances of presenting proofs that follow below are ordered temporally. Still images from the classroom are used to illustrate gestural, diagrammatic, and written modes of communication. Using still images to capture different visual semiotic resources—such as: the appearance of a figure, the stages of a gesture—together with transcriptions of the language—spoken and written—that is used in the situation is a reasonable practice given that my primary aim in transcribing the episodes is to describe how resources from different modalities are used to accomplish the semiotic work of performing a proof in geometry and proofs are, by nature, sequential.

Multimodal description of students presenting proofs

I describe three episodes of students presenting proofs in geometry. The first episode shows student using the written mode to present a proof, in a practice I refer to as *transcribing* a proof. The second episode shows students presenting the written record of the proof by engaging in transcription, and then describing the steps of the proof using the verbal mode and (primarily) indexical gestures to refer to parts of the diagram. The third episode is of a student enacting a proof for the class without transcribing. During this enactment, the student uses a combination of semiotic resources to generate the proof, however the coordination of these resources differs from how they are coordinated by experts[19] engaged in *chalk talk*.

Students Present Proofs I: Transcribing Written Records of Proofs

Episode LV-111303-4P-S3 captures two students presenting proofs to the class in an activity I call *proof transcription:* a proof that the students have previously completed is copied, mark for mark, onto the whiteboard. The episode begins with the teacher, Lucille[20], saying to the class "why don't we put a couple of the proofs on the board" (timecode 0:00-0:09). The proofs were assigned to the students for homework and were completed by the students on worksheets. The canvas for the presentation of the proofs is a whiteboard, and the styluses are wide-tipped whiteboard markers. The teacher calls three students to the whiteboard to present proof problems from the homework worksheet. As the teacher solicits volunteers, she writes (1) and (2) on the whiteboard, to indicate the part of the canvas she wants each of the volunteers to use to present their proofs.

[19] One such expert is the student's teacher.
[20] Names of teachers are pseudonyms.

The two volunteers go the whiteboard. After some initial (inaudible) conversation between themselves, the volunteers begin the work of writing their proofs on the board. Figure 4 shows the two students at the whiteboard at the beginning of their presentations.

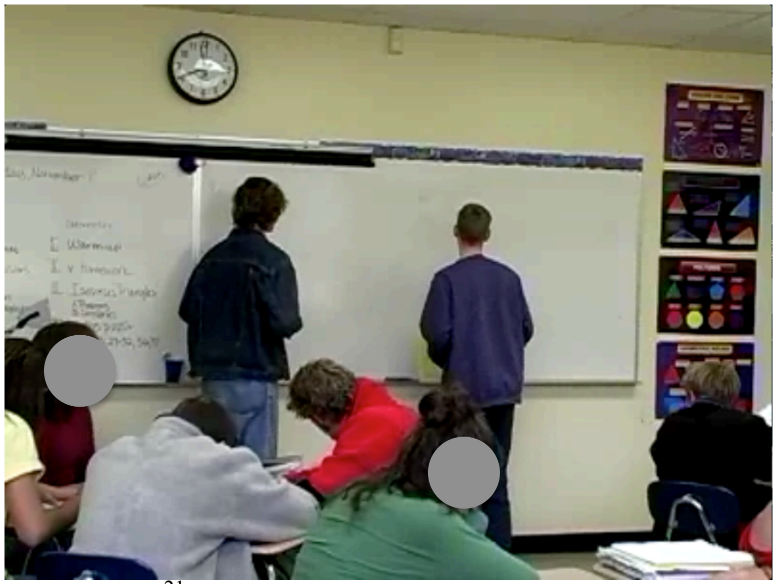

Figure 4: Students[21] at the board presenting proofs (timecode 1:27.20[22])

Each student has with him the worksheets on which they have already completed the proofs they are presenting, resting on the marker tray that is adjoined to the bottom of the white board (yellow piece of paper visible behind the left arm of the student on the right; the student on the left is blocking the view of his worksheet). Figure 4 captures the students after they have had the inaudible preliminary exchange and are getting ready to begin the task of presenting their proofs.

The student on the right in Figure 4 is looking at the space on the whiteboard where he is to present the proof. He holds this gaze for (approximately) sixteen-hundredths of a second second (5 frames) before he shifts his gaze downward to his worksheet. Figure 5 (below) shows the students looking down at their worksheets.

[21] Where necessary, faces are obscured.
[22] Timecodes are specific to video file. Timecode format is minutes:seconds:frame-in-second. The video was captures at 30 frames per second. The number after the decimal ranges from 00-29.

Figure 5: Students looking at their worksheets before presenting proofs (timecode 1:27.29)

This shift in student gaze—from their worksheets to the whiteboard and back—occurs continually throughout their presentations of the proofs. The student on the right holds his initial downward gaze at his worksheet for almost 2 full seconds (approximately 58 frames) before looking back up at the whiteboard and raising his marker to start writing. Once he has raised his marker to start writing, the student once again looks down at his worksheet, this time holding his gaze for approximately .5 seconds (13 frames) before turning his attention back to the whiteboard, then beginning to move his marker. Figure 6 (below) shows the student with his marker raised, ready to start writing, looking down at the worksheet.

Figure 6: Student (on right) with marker raised, consulting worksheet (timecode 1:31.16)

The student on the right has been in position to start presenting the proof for approximately 10 seconds. Once he sets up his worksheet, he looks at the space where he will write the proof on the whiteboard, then he looks down at his worksheet, then looks back to the space where he will write the proof on the board, and finally looks back down at his worksheet. This gazing down and gazing out at the whiteboard occurs before the student makes any inscriptions on the board.

The student on the left can also be seen consulting his worksheet, though the camera angle makes it easier to observe shifts in gaze for the student on the right. The pattern of looking down at the worksheet then looking out at the board and back continues during each student's presentation of the proof. The student on the left is at the board for nearly 2 minutes and 30 seconds. During this time, there are 25 instances—of varying duration—during which he gazes down at his paper.

The student on the right takes nearly 6 minutes to transcribe his proof, and the camera does not capture all of his transcription work at the board. The camera does capture the first two minutes of his transcription, during which the student draws the diagram and writes the given and prove statements on the board. During this time, the student gazes down at his paper 24 times—once every 5 seconds on average—at a rate that is in the same ballpark as the student on the left—who gazes at his paper once every 4 seconds while transcribing the diagram.

That a student would come to the board and transcribe a proof from an already completed worksheet occurs in other geometry classrooms. Figure 7 is a screenshot from a different geometry classroom that shows 3 other students engaged in transcribing proofs.

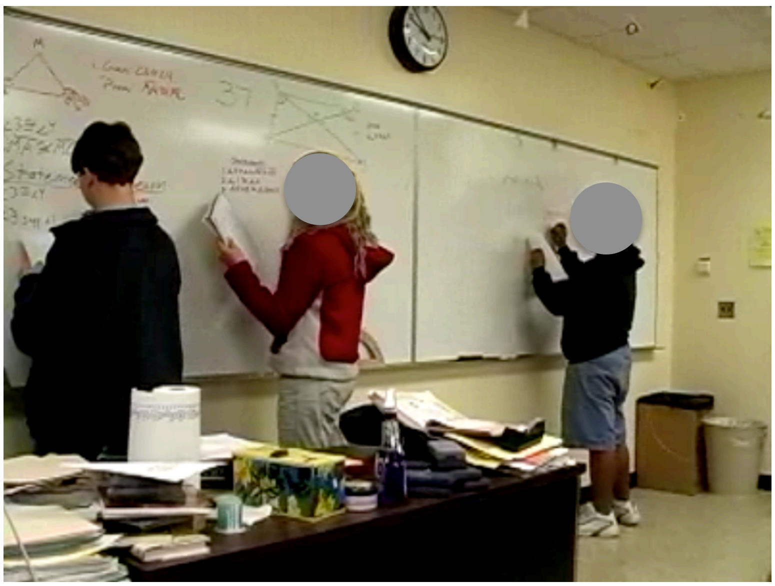

Figure 7: Students in Megan's class transcribing proofs at the board.

In the episode from Lucille's classroom described above, the students are called to the board to present their proofs. The teacher then leads the class in a discussion of each proof, during which Lucille and the students in the class check the details of each proof.

In the episode from Megan's classroom shown in Figure 7 (above), the students initially go to the board to transcribe their proofs, using the same routines as those used by the students in Lucille's class: the original written record of the proof that the students previously generated is on a separate piece of paper that the students continually reference while they are writing the proof on the whiteboard. The task of presenting the proof at the front of the room is one of transcribing inscriptions from their notes, work the students complete in relative silence with their backs to the other students in the classroom. Following the transcription of the proof however, the students in this episode are then asked by the teacher to explain what they did. How students use speech, gesture, and the written record of the proof that they have transcribed onto the whiteboard is discussed in the next section.

In the episodes analyzed above, I described instances of students relying on their worksheets when presenting proofs on the board in geometry classrooms. The regular shifts in gaze from worksheet to whiteboard suggest that the activity of writing the proofs on the board is constrained by the students' visual working memory (Alvarez & Cavanagh, 2004). The segments of the proof that students recreated in between gazes at their worksheets were those they can recall from a glance. The gaze-write-gaze-write pattern of activity described above is evidence that supports this claim. How the students used the worksheets suggests that the task for the students when presenting proofs was one of visual reproduction. The students in the episodes described above worked in virtual silence, with their backs facing the classroom, and appeared to be under no expectation to provide an account of what they are doing as they completed the reproduction. This practice of *proof transcription* is a marked departure from the disciplinary practice of *chalk talk.*

Students Present Proofs II: Explaining a Written Record of a Proof

In episode MK-111102-3P-S2, after the students have transcribed their proofs, they are asked by the teacher to explain them. The student that has completed number 38—the student who is in the far right of Figure 7, above—is called on to explain his proof first. The student begins by saying: "Uh [pause] yeah [pause] alright. The given for 38 was angle 1 is congruent to angle 4, which are the two base angles of triangle N E T" (timecode 5:46 – 5:52). Figure 8 shows a reproduction of the diagram that the student's proof refers to.

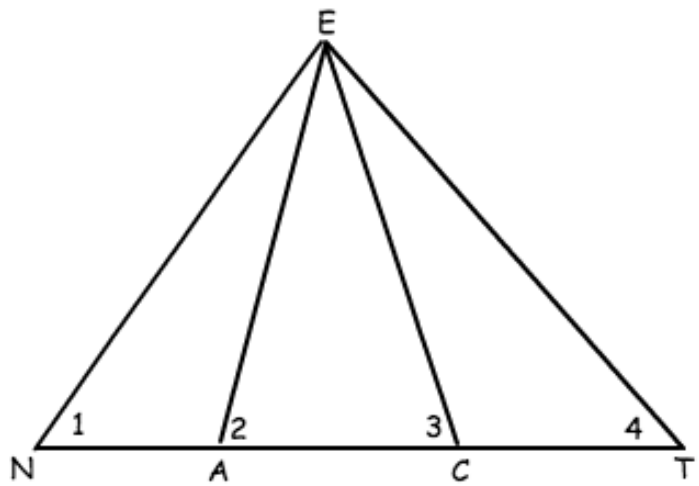

Figure 8: Diagram showing triangle NET

As the student says "two base angles of triangle N E T", he points first at angle 1, then at angle 4, as shown in Figure 9 and Figure 10 below.

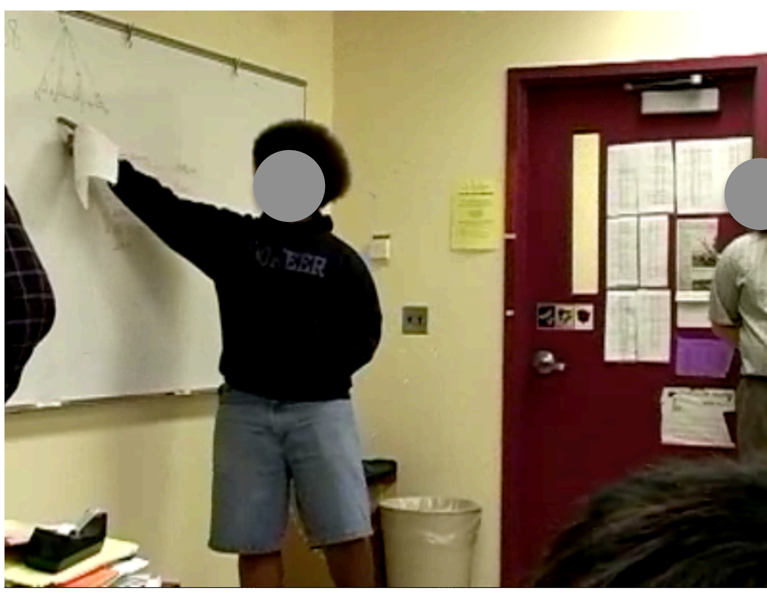

Figure 9: Student at the start of the indexical gesture that coincides with "two base angles". Here, the student is pointing at angle 1 (timecode 5:53.07)

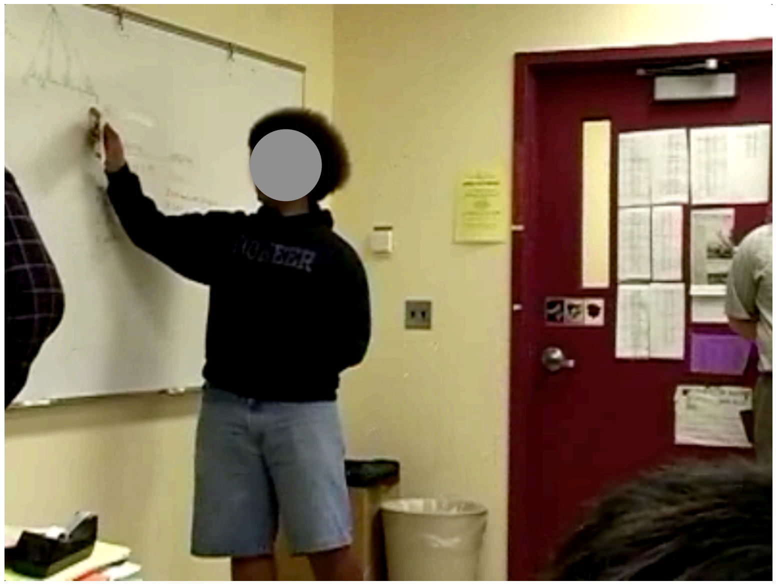

Figure 10: Student at the end of the indexical gesture that coincides with “two base angles”. Here, the student is pointing at angle 2. (timecode 5:53.16)

The student begins his explanation by stating the givens for problem 38. He initially identifies the angles using their diagrammatic labels (i.e., “angle 1” and “angle 4”) and then the angles are described conceptually as the “base angles” of the given triangle. The conceptual description is accompanied by a pointing gesture (shown above). Once the student has established the givens, he proceeds through a step-by-step narration of the written record of the proof.

During this narration, the student uses speech to state what is already written in the statements and reasons of the proof, with almost no deviation with the exception of adding some verbal cues for pausing or transitioning. The student holds his worksheet in his right hand throughout the narration, and the student uses his right hand to gesture at the diagram as he speaks. His gestures during the verbal recount of the written record of the proof are exclusively pointing (indexical) gestures that indicate the diagrammatic referents of the steps he is describing.

Figure 11 shows a combined transcript of the student’s verbal recount of the proof with the written record of the proof that is transcribed onto the whiteboard. The verbal recount of each step appears below the written record of the proof, in italics. Moments during the recount where the student makes indexical gestures are marked by asterisks (*)

<table>
<tr><th>Statements</th><th>Reasons</th></tr>
<tr><td>$\angle 1 \cong \angle 4$ written as $<1 \cong <4$, $\overline{NA} \cong \overline{TC}$</td><td>Given</td></tr>
<tr><td colspan="2">The given for 38 was angle 1 is congruent to angle 4, which are the two base angles (*) of N E T, and then N A is congruent to T C. So, from there</td></tr>
<tr><td>$\overline{EN} \cong \overline{ET}$</td><td>2 ≅ base <’s→2≅ opp. sides</td></tr>
<tr><td colspan="2">I put E N is equal to E T because two congruent base angles gives you two congruent opposite sides, which would be these two (*). And then from there</td></tr>
<tr><td>△ENA ≅ △ETC</td><td>SAS post.</td></tr>
<tr><td colspan="2">You can say triangle E N A is congruent to triangle E T C, because of the side angle side postulate, right there (*). And then, from there, you can say uh</td></tr>
<tr><td>$\overline{EA} \cong \overline{EC}$</td><td>CPCTC</td></tr>
<tr><td colspan="2">E A is equal to E C, because of CPCTC. And then uh</td></tr>
<tr><td>$<2 \cong <3$</td><td>2≅ opp. sides→2≅base <’s</td></tr>
<tr><td colspan="2">These two angles (*), angle 2 and angle 3, are congruent because two congruent opposite sides gives you two congruent base angles.</td></tr>
</table>

Figure 11: Combined transcript of written proof and verbal/gestural recount of the proof.

One interpretation of the student’s verbal explanation of the proof is that it is a step-by-step account of the written steps of the proof. The student uses the past-tense almost exclusively to describe the work, indicating that, for the student, the proof is something that has happened already. During the recount, the student uses conceptual language (e.g., base angles, opposite sides) and deictic gestures to indicate specific parts of the diagram that are referenced directly in the written statements of the proof—e.g., describing angles 1 and 4 as *base angles*, pointing to segments EN and ET. The student’s use of speech adds little to the proof that is not communicated already through its written statements or reasons and accompanying diagram. Likewise, the student’s use of indexical gestures serves to re-iterate the links between the written steps of the proof and the objects in the diagram that are realized already by the fact that the written statements are linked to the diagram. Thus, there is an overlap in the work done by the spoken, written, and gestural modes of communication.

It is the case that the gestures that the student uses to point to different parts of the diagram anchor indexicals in the student's speech to specific parts of the picture. However, in every instance where the student uses indexical speech, the student also uses diagrammatic referents—such as the labels of angles or segments—to name the parts of the figure that the statements he is making are about. The gestures are necessary to clearly indicate the referents of the indexical speech, however the indexical speech itself is auxiliary to the diagrammatic referents the student uses in every step of the proof.

Using speech to provide a step-by-step description of the written record of the proof is evident in another student's explanation of her proof. This student—shown working on the proof in the middle of the whiteboard in Figure 7, above—is also called to explain the proof once she has finished transcribing. She uses the verbal and gestural modes to provide the explanation in similar ways to her classmate (described above). Figure 12 shows a recreation of the diagram the student used when writing the proof.

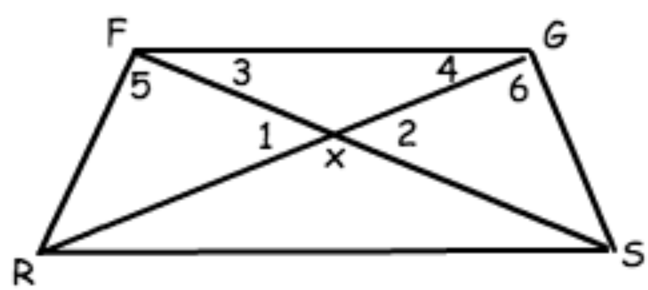

Figure 12: Recreation of the diagram the second student used to write the proof.

Figure 13 displays a combined transcript of the written record of the proof and the student's step-by-step verbal explanation of it. The student's speech is displayed in italics underneath the written statements and reasons of the proof. Teacher interjections are boldface type between square brackets. Student indexical gestures are marked in the transcript by asterisks (*), and moments where the student adds markings to the diagram during the explanation are indicated by (**m**).

<table>
<tr><th>Statements</th><th>Reasons</th></tr>
<tr><td><5 ≅ <6, $\overline{FR} \cong \overline{GS}$</td><td>Given</td></tr>
<tr><td colspan="2">But so um…they tell you that angle 5 (*) and angle 6 (*) are equal and they tell you that FR and GS are equal. And</td></tr>
<tr><td><1≅ <2</td><td>Vertical <’s ≅</td></tr>
<tr><td colspan="2">You know that angle 1(m) is equal to angle 2 (m) because of vertical angles, they give you that. And then</td></tr>
<tr><td>$\triangle FXR \cong \triangle GXS$</td><td>AAS</td></tr>
<tr><td colspan="2">Because of, um, AAS, you know that angle FXR and GXR are congruent— [teacher says: “triangles”]—triangles, sorry. And um</td></tr>
<tr><td>$\overline{FX} \cong \overline{GX}$</td><td>CPCTC</td></tr>
<tr><td colspan="2">Then you know that FX is congruent to GX because of CPCTC. These two (m)—[teacher says: “put two marks on them so we know they are not equal to the other ones”]—yup. (m) And</td></tr>
<tr><td>$\triangle FXG$</td><td>Def. Isos. $\triangle$</td></tr>
<tr><td colspan="2">So you know that FXG (***) is an isosceles triangle cause they’re equal. And then ah</td></tr>
<tr><td><4 ≅ <3</td><td>Isos. $\triangle$’s →≅base <’s</td></tr>
<tr><td colspan="2">4 is congruent to 3 cause isosceles triangles give congruent base angles</td></tr>
</table>

Figure 13: Combined transcript of student written record and verbal/gestural recount of a proof.

Unlike the example considered above, the narration of the proof transcribed in Figure 13 unfolds primarily in the present tense. As the student describes the proof, she makes pointing gestures at the diagram and adds markings to indicate parts of the figure that are related to each other. Figure 14 – Figure 22 (below) show still images of the student gesturing toward and marking the diagram. The figures are listed in the order in which the student makes the gesture or marking in the steps transcribed above. To point to triangle FXG, the student points to each of its vertices. I represent this in the transcript with three consecutive asterisks. These gestures are illustrated in Figure 20 - Figure 22.

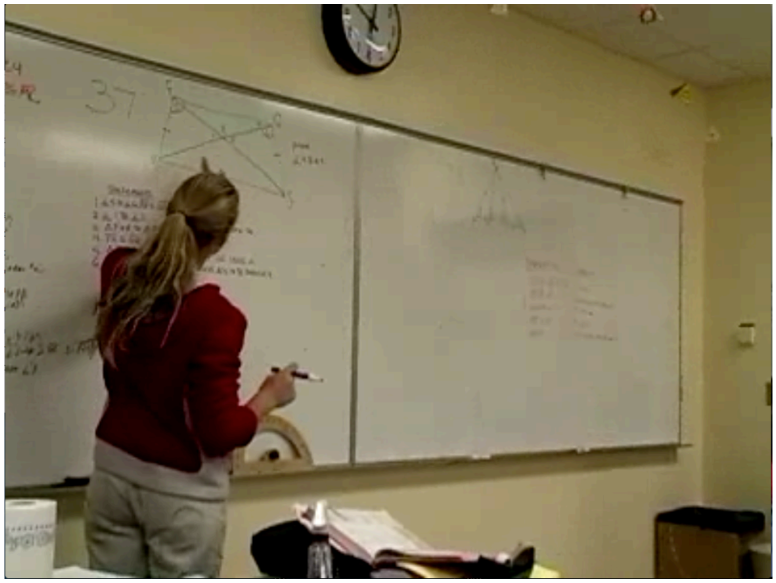

Figure 14: Student pointing while saying “angle 5.”

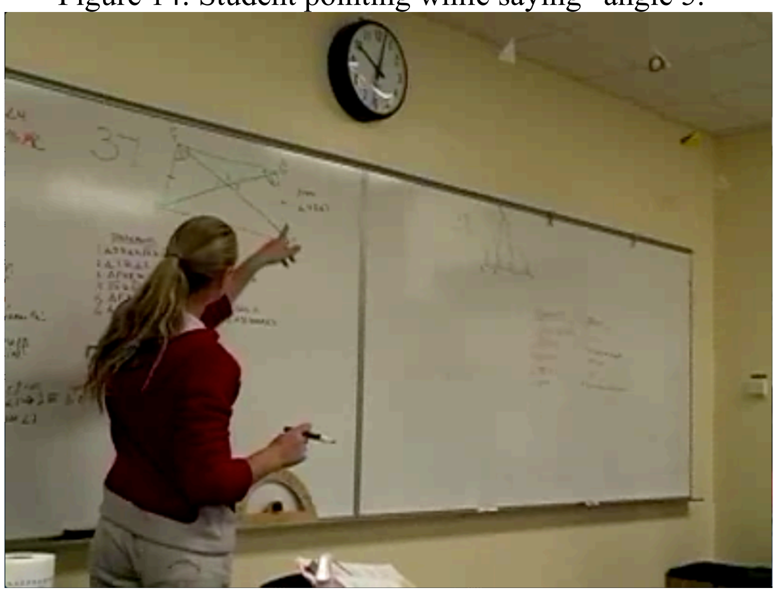

Figure 15: Student pointing while saying “angle 6.”

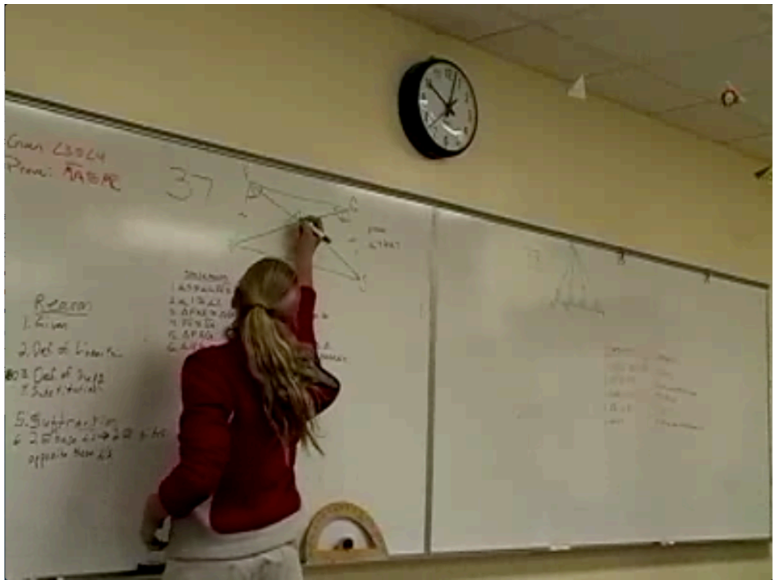

Figure 16: Student adding congruence mark to angle 1.

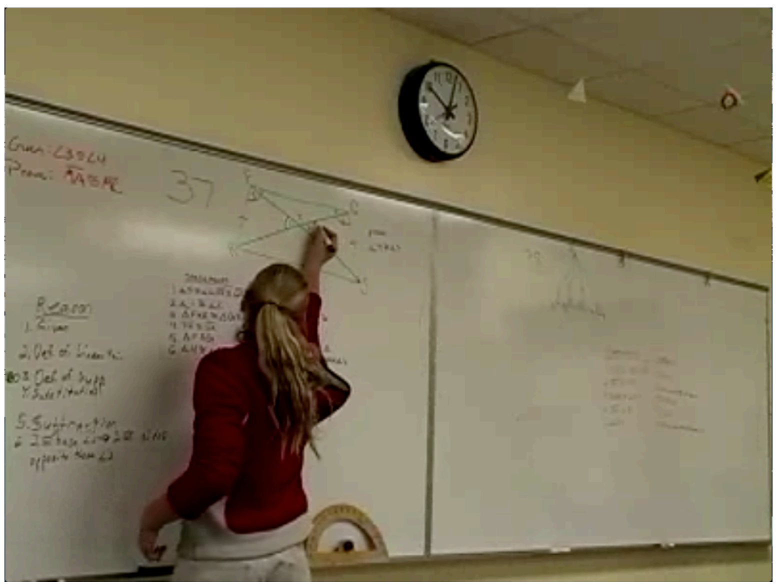

Figure 17: Student adding a congruence mark to angle 2.

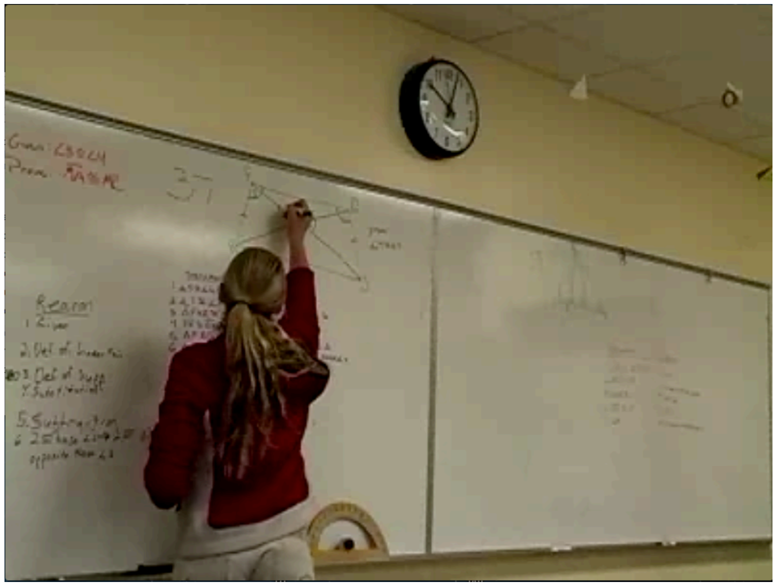

Figure 18: Student adding a congruence mark to $\overline{FX}$

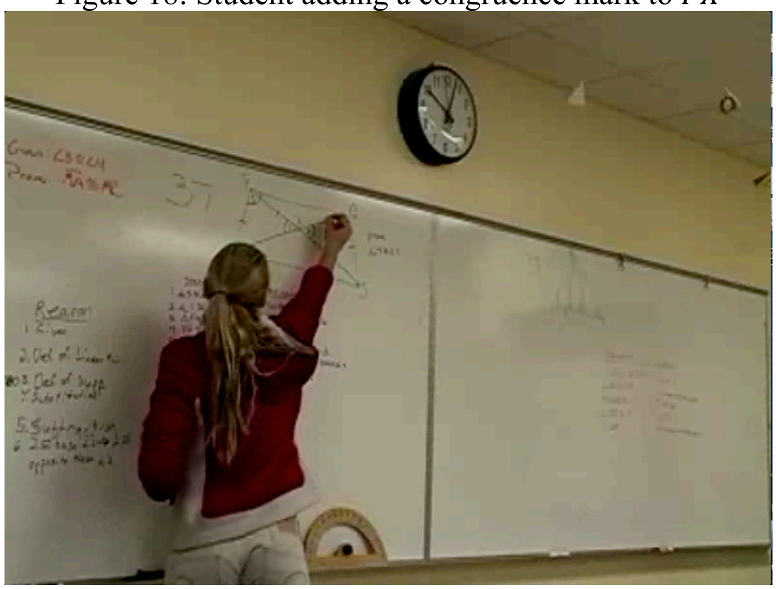

Figure 19: Student adding a congruence mark to $\overline{GX}$

Figure 20: Student pointing at vertex F

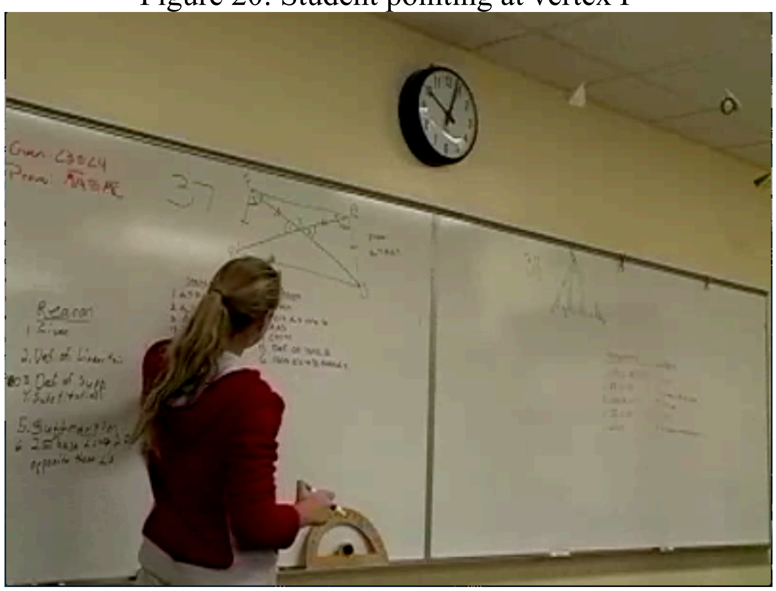

Figure 21: Student pointing at vertex X

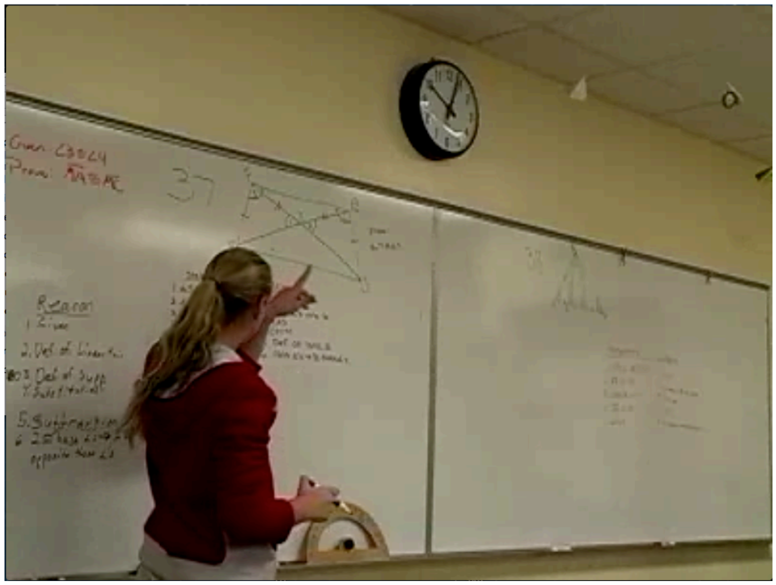

Figure 22: Student pointing at vertex G

Like her classmate, this student's verbal explanation of the proof is essentially a spoken translation of the statements and reasons she has written on the board, with a key difference being that this student describes the steps in the present tense—i.e., as steps that are taking place—as opposed to the past tense used by the previous student.

The student's use of the present tense is coordinated with the student adding markings to the diagram as she describes the steps in the proof to indicate the relationships that are established as the steps of the proof develop. The gestures that the student makes at the diagram are used to point at the different features of the diagram as she speaks. Beyond gesturing at the diagram, this student also adds congruence markings to angles and segments that she establishes as being congruent. In both instances, the move seems to combine a deictic gesture—*these are the angles I am talking about*—with the application of a diagrammatic attribute (Dimmel & Herbst, 2015) to indicate congruence. In this case, the attributes that are added to the diagram are small arcs that indicate the angles are congruent. The deictic role of adding the markings is highlighted when she prefaces adding markings to the congruent segments with: "these two".

The above examples of student explanations of previously transcribed proofs show students using speech redundantly to translate information that is already established in the written record of the proof. Unlike when mathematical experts present proofs via *chalk talk*—during which a mathematical expert generates a mathematical work in real time at a canvas—when students are called to present proofs, the focus of the task is on achieving an accurate transcription of a previously completed proof onto the whiteboard. Once the transcription task has been completed, a satisfactory explanation of the proof is a step-by-step translation into speech of what is already established in the written record of the proof.

In the first case, the student used the past-tense to recount the proof. In the second case, the student used primarily the present-tense and augmented the verbal explanation by adding markings to the diagram. Because of the use of the present tense and the addition of markings to the diagram that is coordinated with the verbal development of the proof, the second student's presentation could be seen as a closer approximation to the disciplinary communication practices that are used by mathematical experts when presenting proofs via *chalk talk.*

The teacher in the video does not comment on these differences in the student presentations. I call attention to this to point out that the differences in how the students use modes of communication to present the proof would be an opportunity for the typical work that students do in geometry classrooms to be connected to disciplinary-specific communication practices. The fact that the teacher does not comment on how the students used the modes of communication to present the proofs suggests that what each student did was equally acceptable. It also suggests that the teacher is not focused on the meaningful differences in these ways of presenting a proof.

In each of the examples considered above, the proof that was presented was completed already by the students that were called to the board to share their work. There are situations where a student might be presenting a proof to the class without having completed the proof ahead of time. In such situations where a student is at the whiteboard generating the proof—with the help of the teacher and the class—how are the various modes used by the student to present the proof? The example analyzed below considers such a circumstance.

Students Present Proofs III: Generating a Written Record of a Proof

Episode MK-120502-3P-S7 begins with the teacher setting out to prove that a quadrilateral that has one pair of congruent, opposite sides is a parallelogram. The teacher begins the presentation of the proof by drawing and labeling a diagram of a quadrilateral. The quadrilateral is drawn to look like a parallelogram. About the choice to draw the diagram to look like a parallelogram, the teacher, Megan, says: "I'm gonna draw it to look like a parallelogram, because your book is gonna, um, do that most of the time." (timecode 0:09 – 0:15). The teacher goes on to write the given and prove statement by soliciting responses from students in the class. The teacher says to the class: "one pair of sides is congruent, and it's parallel." She tells a student to: "pick a pair of opposite sides". The student volunteers "AD" and "BC". The teacher says: "we'll say that AD is congruent to BC" and that: "it's parallel. AD is parallel to BC". As she speaks, the teacher also writes these statements next to the word "Given" underneath the diagram (see Figure 23).

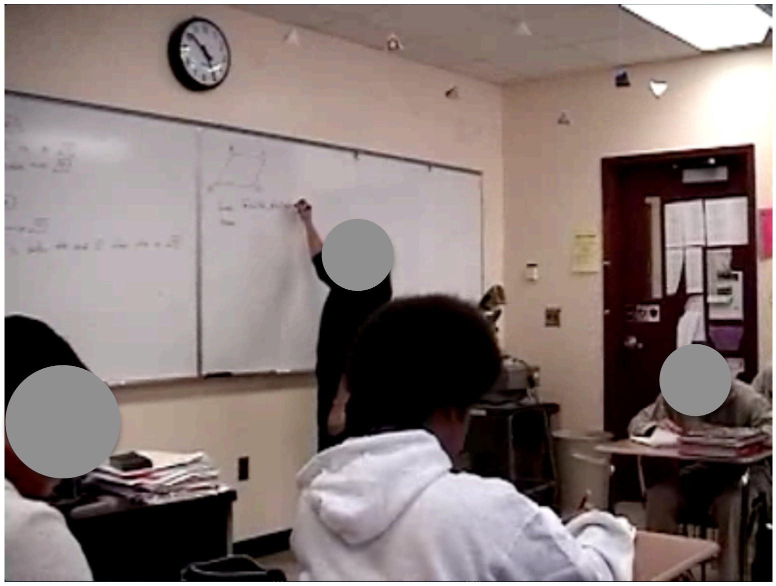

Figure 23: Teacher writing the given statement while speaking

The teacher goes on to say: “OK. I want to prove that this thing is a parallelogram. Prove [writing as she speaks] ABCD is a parallelogram” (timecode 1:00 – 1:09). The teacher then writes the words “Statements” and “Reasons” (underlining each) to the right of the diagram. She says: “Ok. Let’s put my given in”, then reiterates the statement that is given as she writes it next to the numeral (1) under the statements heading. Having written the givens as step (1), the teacher asks the students in the class if “anyone has an idea” for how to get started. A student volunteers, and the teacher tells the student to “get up here and start writing.” In this circumstance, the student has been called to the board to present a proof to the class that he has not written out ahead of time. Here is an opportunity for the teacher in the room to guide the student presentation of the proof. The student accepts a marker from the teacher and approaches the board without anything else in his hands.

The student begins the presentation by saying, “I just thought it might be easier if we made it into two different triangles”. The student uses speech to suggest an overall strategy for doing the proof and to provide a rationale for that strategy–he thinks it “might be easier”. This use of speech is consistent with how mathematical experts use speech during *chalk talk* to make explicit the strategic value of what they are doing as they do it (Artemeva & Fox, 2011; Greiffenhagen, 2014). The student's use of speech in this way could also be read as an indirect request for permission to add an auxiliary line to the diagram. Adding such a line is not an action that students normatively take when doing proofs in geometry (Herbst et al., 2009). Once he makes this suggestion, the teacher says, "go ahead, put one in", thus giving the student permission to add an auxiliary line to the diagram. The student then draws in the line, as shown in Figure 24.

Figure 24: Student adding auxiliary line to the diagram

After drawing in the auxiliary line—segment AC—the student says: "We'd have that, uh, AC is congruent to AC" (timecode 2:39 – 2:42). After the student finishes saying that AC would be congruent to AC, he moves to the "Statements" column and writes this statement next to the numeral (2). Next to the numeral (2) in the "Reasons" column, the student writes the word "reflexive". While the student is writing the word "reflexive"—when he is almost finished writing the word—the teacher asks: "Why is it congruent"? The student finishes writing the word and then says: “reflexive.” In this segment, the student’s use of the spoken and written modes occur in sequence. The student does not write *as* he speaks, but rather speaks, then writes; or: writes, and then speaks.

The teacher continues to help the student at the board to generate the proof by prompting him to identify a “little subgoal” that will help them get to the main result they are trying to prove. This move by the teacher is an attempt to prompt the student to describe the overall plan of the proof before listing the specific statements and reasons the final proof will comprise. With continued prompting from the teacher and some assistance from other students in the class, the student identifies that he is trying to prove that the triangles are congruent, because then he would be able to show that the quadrilateral has two pairs of congruent opposite sides, and would therefore be a parallelogram[23]. The teacher asks the student at the board how he plans to show that the triangles are congruent, and he says “alternate interior angles”. The teacher says: “yeah, show me some”. At which point, the student goes to the diagram and adds numeric labels (1, 2, 3, and 4) to the angles that he is referring to. This is another instance of using the addition of diagrammatic attributes—in this case, numeric labels for angles—deictically to identify a referent in the diagram that the student is making claims about. Figure 25 shows the student adding labels to one of the alternate interior angles.

[23] This portion of the episode is paraphrased, rather than transcribed, because it unfolds through collective/overlapping utterance from several sources that are difficult to parse and represent as distinct turns of speech. What I have captured in the paraphrase is the result of this interaction between the student presenting the proof at the board, the teacher, and the other students in the class.

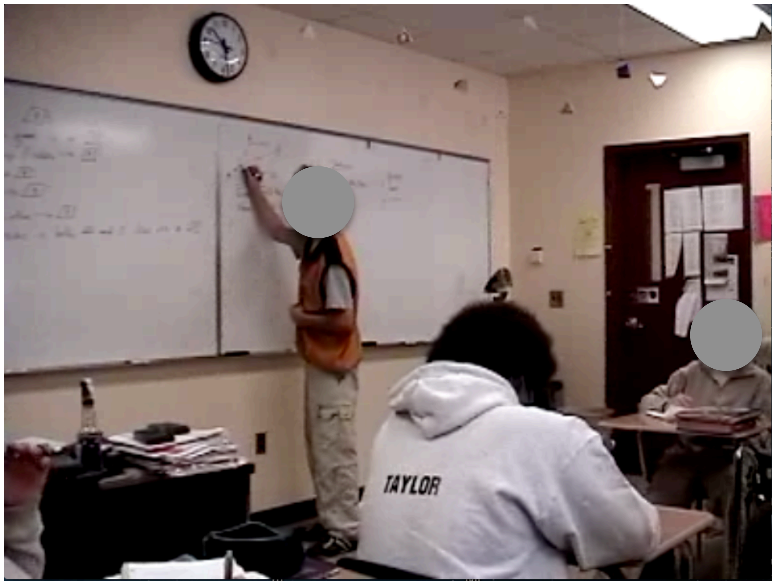

Figure 25: Student "pointing" at an angle in the diagram by labeling it.

During this portion of the presentation, the teacher's assistance can be seen as providing a scaffold to help the student execute the communication practices that a disciplinary expert might use to present a proof to the class. The teacher is helping the student to use the verbal mode to describe the overall plan for the proof, while also using diagrammatic markings to identify relationships in the diagram. The teacher does not state—for the benefit of the other students in the class—that what she is doing is helping the student coordinate the different modes of mathematical communication into a presentation that is similar to those that mathematical experts make when they are presenting proofs, but she none the less helps the student model approximations to these practices.

Once the student labels the angles, the teacher says: "Ok [student name] which ones are congruent?" . The student says: "angle 1 is congruent to angle 2, and angle 3 is congruent to angle 4". At this point, it is not yet known that angle 3 and angle 4 are congruent. The teacher catches this and asks the class: "Ok, which ones is he not correct?". Several students chime in[24]

[24] These different contributions are difficult to hear distinctly and are not transcribed.

to indicate that “angle 3 and angle 4” are not yet known to be congruent. As these other students in the class speak, the students at the board says: “Oh, right”. The teacher asks the student at the board: “do you know why?”. The student at the board replies by saying: “I don’t know if these [puts his finger on one side of the parallelogram and subtly taps it as he says “these”] are parallel”. The teacher instructs the student to “put the ones you do know” to be congruent. The student then uses these labels to write a statement of angle congruence next to the numeral (3) under the statements column. The student writes the statement—“$\angle 1 \cong \angle 2$”—and reason—alternate interior angles theorem—without speaking, just as he did for the statement and reason written under step (2). Once he identifies that these angles are congruent, he marks them as congruent in the diagram. He adds the congruence markings to the diagram without speaking.

After the student adds the congruence markings to angle 1 and angle 2, the teacher asks: “OK. What do you have?” The student says: “Side Angle Side”. The teacher repeats: “Side Angle Side”. Then, in silence, the student writes the statement (step 4) that: “$\triangle ABC \cong \triangle CDA$” with “SAS” as the reason. The teacher asks the student: “Then what?”. The student says: “Then we have that BA is congruent to CD”—teacher interjects “right”—then the student says: “CPCTC”. The student then writes this step in silence and adds congruence markings (three hash marks) to segments BA and CD. The teacher says: “Ok. Then?” The student says: “Then you have that by/you have the definition of parallelogram—[25]”. The teacher interjects: “No I don’t. What do I have? Someone help him”. At which point, a class discussion ensues to clarify the warrant they have for establishing that ABCD is a parallelogram. A different student in the class says that “two pairs of congruent opposite sides means it’s a parallelogram”, and the teacher tells

[25] The “/” indicates a point where the student interrupts himself; the “—” indicates a point where the student is interrupted by the teacher.

the student at the board to “start writing”. The student then finishes writing this step in the proof in silence.

Review of student presentations of proofs

The students in the episodes described above tended to write (i.e., transcribe) proofs before verbalizing their steps. The gestures that students used while speaking were primarily indexical. The writing that they did while speaking was limited to adding markings to the diagram. In addition to using different modes of producing semiotic resources in sequence—rather than in parallel—students used resources from different modes redundantly, rather than expansively—e.g., using speech to vocalize the statements and reasons that are written in the steps of the proof; using deictic gestures to point to objects that are identified using diagrammatic referents. The episodes analyzed above suggest that teachers primarily expect students to reproduce accurate written records of (already completed) proofs when presenting proofs to the class.

Teachers continually correct students in mathematics classrooms. In the examples of checking proofs that are considered below, there are examples of teachers adding what might be described as very minute details to student proofs. Yet, during the video episodes analyzed above, teachers were not correcting or guiding or advising students on how to use semiotic resources when presenting proofs. Part of this is likely on account of the fact that, as mathematical experts, teachers are likely to take their literacies in discipline-specific communication practices for granted (Kress & van Leeuwen, 2006; Kress, 2003). They might expect that as students become more accomplished at doing proofs, their presentations of proofs will begin to approach the level of explanation that teachers provide when they are presenting proofs. But there is a growing body of literature that indicates that people can be taught explicitly about coordinating semiotic resources from different modes—e.g., getting guidance about how to gesture when explaining a concept—and that, following such instruction, people's literacy in that mode of communication improves (Alibali et al., 2013; Goldin-Meadow & Wagner, 2005; Yore, Pimm, & Tuan, 2007). Such work suggests that teachers might affect how students learn to communicate about proof directly by talking about that communication—e.g., raise your voice to emphasize the key concepts, describe orally what you are writing with spoken words as you write it. That the multimodal aspects of how a proof is presented were not—in the episodes considered above—the subject of teacher intervention suggests that geometry teachers have an expectation that "presenting a proof" means presenting its written steps, one-by-one. How this expectation can be described as a semiotic norm of the situation is considered below, following the descriptions of how the details of proofs are checked in geometry classrooms.

Descriptions of episodes of *checking proofs*

The analysis above provides an account of how teachers and students use the spoken, written, gesturing, and diagramming modes to present proofs to the class. An activity that can follow the presentation of a proof is the *checking* of a proof. When a proof is checked by the teacher (and the students in the class), its details are scrutinized to the end of making sure that it is a complete, accurate proof of what was to be shown. The lesson corpus contains instances of proofs being checked by the teacher and the students in the class. In this section, I describe episodes during which the details of proofs are checked in geometry classrooms. The purpose of analyzing the episodes is to formulate conjectures about the normative ways in which the details of proofs are checked.

In each episode in the corpus that features a student presentation of a proof (or: a presentation of a student proof, as will be shown below), the completed proof, once presented, is subjected to scrutiny by the teacher and students in the class. The activity of scrutinizing the proof begins with the teacher asking the class an open-ended question whose target is the presented proof, such as: "Any thoughts on this proof?" Following this question, there could be a period of silence while the students in the class review the proof. After a pause of 5-30 seconds (or even longer), students in the class make comments about the aspects of the proof that require further elaboration, clarification, or correction. Below, I describe four instances of the details of the proof being checked by the teachers or students in the class.

Checking proofs (I):midpoints and segments.

One instance of the details of a proof being scrutinized occurs in episode LV-111303-4P-S4. In this lesson, students have written proofs on a classroom chalkboard, in accordance with the proof-transcription routines that are described above. The teacher is now reviewing the proofs with the students in the class. During the review, there is an exchange between the teacher and a student about whether it is necessary to include an explicit written step that establishes the congruence of two midpoint segments. The givens of the problem include that the midpoint of segment JA is L. The proof that is written on the board includes a step that establishes the congruence of segments JL and LA, by the definition of midpoint. Figure 26 shows a screenshot of the proof.

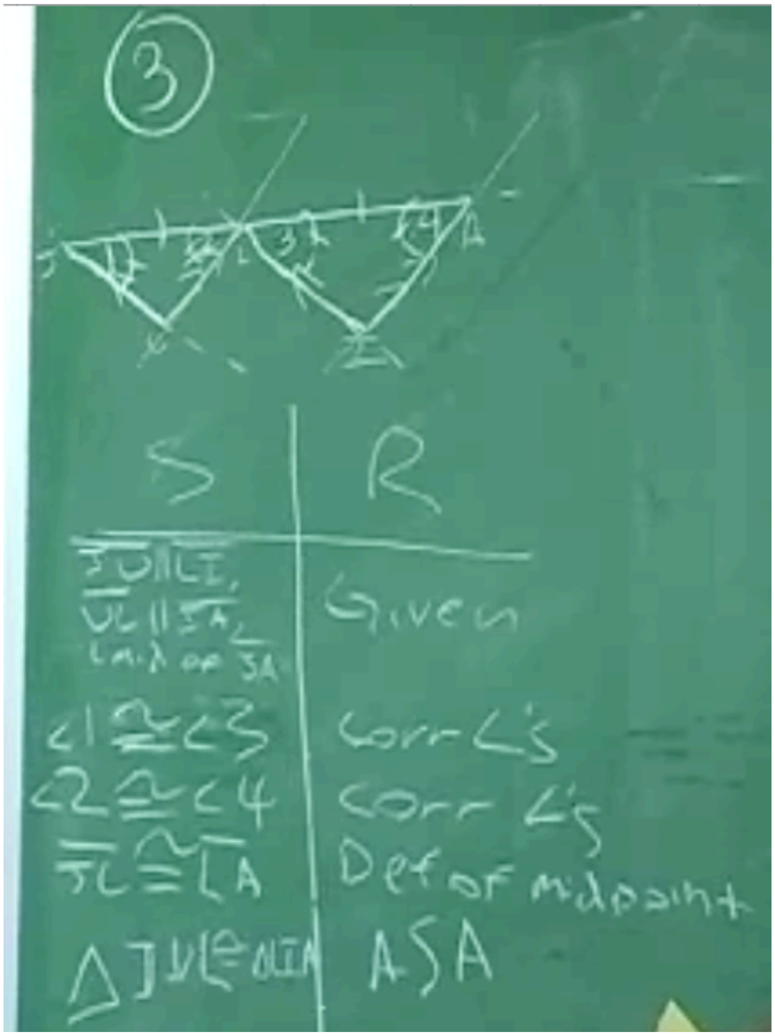

Figure 26: A proof that is being checked in a geometry classroom.

A student asks the teacher: "What if I skipped [inaudible]" then gestures from the given to the statement that the triangles are congruent by ASA. While the student's comment and accompanying gesture are somewhat difficult to parse, helpfully the teacher uses the chalk to make a stroke that at least partially records what the student asked (see Figure 27).

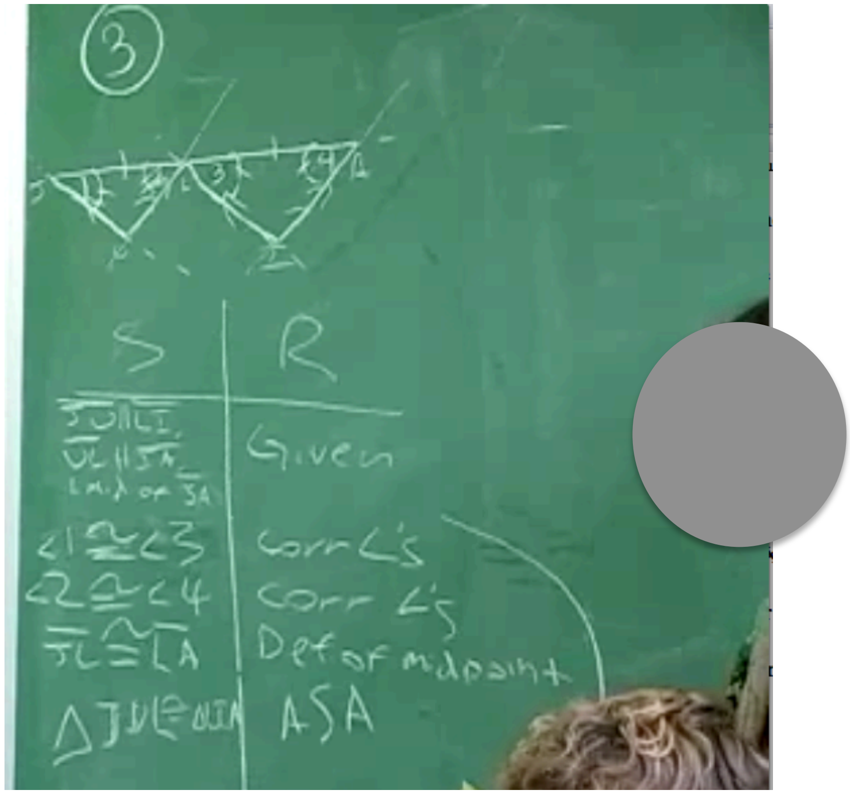

Figure 27: Teacher uses a stroke to represent a gesture accompanying a student question.

Following the student gesture, the teacher says: “This”, then adds the stroke shown in Figure 27. The teacher replies with: “Then you’ve left out a big step.” The student’s participation in the exchange is not audible, however from the teacher’s replies it seems that the student is raising questions about what can be immediately concluded from the information that is given. The teacher says: “We have to say, by definition of midpoint, we know that this [uses a “V” indexical—i.e., the teacher uses her index and middle fingers to form a “V” that she uses for pointing—to point to each part of the segment] is split into two equal parts.” The teacher continues: “And we have to say, because those are parallel [uses a slanting hand gesture to highlight each parallel segment] and this is the transversal [uses a horizontal hand gesture that starts with middle finger just to the left of point A and continues until the back of her hand is at point L], the corresponding angles are congruent.”

The teacher is calling attention to the need to unpack what is entailed by the givens into explicit written statements. The final point that the teacher makes about the level of detail that is necessary for the proof is that, "we realize we are accepting somewhat abbreviated reasons, however, if you went to like another class even in this building, they might want more specific" As the teacher makes this comment, she indicates the reasons—by using a pointing gesture—that say simply "corr. ∠'s". The next episode shows another instance of a geometry teacher helping to make clear to the students what kinds of details are expected (by the teacher) to be included in a proof.

Checking proofs (II): distinguishing objects from properties

Another issue that arises when proofs are being checked is whether the written statements of a proof distinguish objects—such as angles—from their properties—such as their measures. This detail is at issue in episode CM-101802-5P-S4. The teacher leads the class in checking a proof that hinges on a triangle angle sum argument. As a warrant for the claim that the measures of the angles of a triangle sum to 180, a student writes: "All Δ add up to 180 (drawing)". The teacher keys in on this statement of the warrant, saying: "There is one thing I want to say when you write out that theorem. It really is not terribly good form to say um the triangle adds up to one-eighty. It's the angles/the **measures** of the angles that add up to one-eighty. The sum of the measures of the angles equals one-eighty." [4:05 – 4:23, boldface added to indicate rising pitch and volume on the word "measure"] Complementarily, in an episode where the teacher is doing a proof that calls for an angle addition argument, the teacher consistently refers to the *measures* of the angles, both through speech as well as by writing "*m*∠ [x][26]"For example, a teacher would write "*m*∠1" to refer the "measure of angle 1".

[26] Here [x] is a placeholder that refers to the label of the specific angle.

Checking proofs (III): distinguishing definitions from theorems.

Another instance of details that are scrutinized is distinguishing definitions from theorems. In episode CM-101802-5P-S4, there is a discussion about the statements and warrants that are necessary to establish that two angles that form a linear pair are supplementary. The students are given a diagram of a triangle and one of its exterior angles; the angles of the triangle are numbered 2, 3, 4, and the exterior angle is numbered 1. Figure 28 shows a screenshot of the proof.

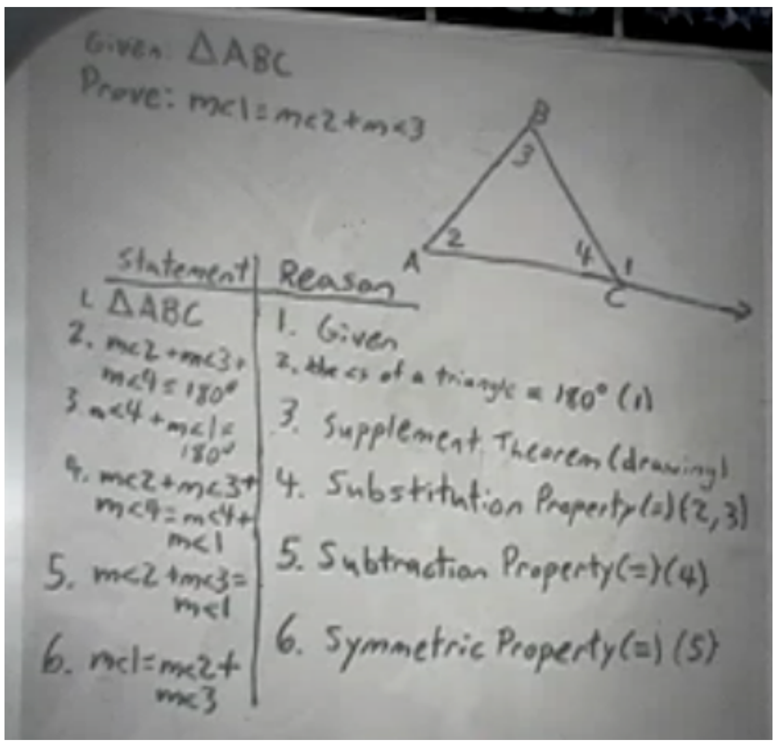

Figure 28: A proof from episode CM-101802-5P-S4.

The teacher presents the proof on an overhead projector. The initial presentation of the proof is followed by roughly 25 seconds of silence [0:15 – 0:40], during which the teacher and the students in the class scrutinize the written details of the proof. The teacher breaks this silence by acknowledging a student who has her hand raised. The student says, “I think on step 3 you have to state that 1 and 4 [pause] form a linear pair before you do that.” The student is keying in on a detail of the written record of the proof that she thinks requires unpacking into more basic steps.

The teacher replies, "OK. If you say 1 and 4 form a linear pair [pause] then can you use the supplement theorem to say they add up to 180?" To which several students in the class respond with "Uh-huh". The teacher calls on a different student, who says: "I just think that you can, uh, say that angle 4 and angle 1 are supplementary before/because you can see that they form a linear pair from the drawing so [pause] you can use the supplement theorem but you have to say they are supplementary before you say they equal one-eighty." To this, the teacher replies with "OK" then calls on a different student. The third student comments on the evidence that is cited in the warrant for step 2. The student says: "For step 2, the reference is just [step] 1 [pause] you could say the drawing." The teacher replies with: "You could say the drawing or the triangle."

At this point, the teacher decides to return to the step that is warranted by the definition of linear pair. It appears that the teacher wants to clearly distinguish what can be deduced from a definition—that two angles form a linear pair—and what requires the application of a theorem—in this case, the target theorem is the *supplements theorem*: "if two angles form a linear pair, then they are supplementary". The exchanges that follow between the teacher and the class illustrate that drawing the distinction between the definition and the theorem is a central point for the teacher.

The teacher says: "I want to get back to the supplement theorem. Can anyone give me the supplement theorem in if-then form so we know what it says?" A student responds to the teacher's question with: "If it forms a/if two angles form a linear pair, then they are supplementary." The teacher says: "OK. Does/does it/it doesn't say one-eighty [pause] it says if they form a linear pair, then they are supplementary." The teacher is directing students to distinguish *a theorem* that allows one to conclude that two angles are supplementary from *the definition* of supplementary angles. The teacher continues along these lines: "So would/we'd have to use the supplement theorem to first say they are supplementary. OK. And then how would we get them to add up to one-eighty?" At this point, the teacher calls on a student who says something inaudible that the teacher repeats: "Definition of supplementary angles. OK. This is something we have been through quite a few times. I think it even came up on the test."

Proceeding from linear pairs to supplementary angles to the sum of the measures of the angles being 180 degrees is a point of emphasis in an instance of proof checking in an episode in Megan's classroom. In a way that is nearly identical to how the scene plays out in Cecilia's classroom, a teacher is scrutinizing the details of a proof that is being presented to the class. The student that presented the proof is at the whiteboard to explain the proof in a manner that is consistent with the student proof presentation routines described above. Figure 29 shows a screenshot of the student's proof.

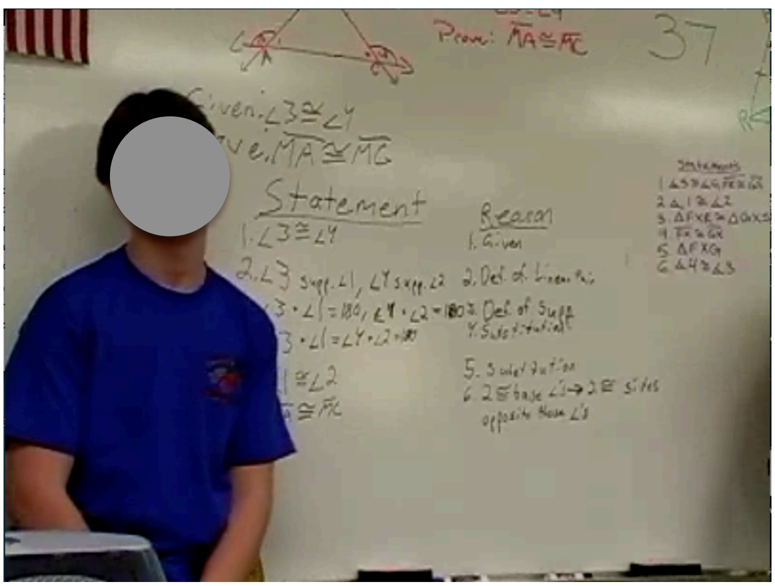

Figure 29: Proof that uses "Definition of Linear Pair" (step 2) to warrant statement that pairs of angles are supplementary.

In the proof, the student states—in step 2—that two pairs of angles—∠1 and ∠3, and ∠2 and ∠4—are supplementary. This statement is warranted by the definition of linear pair. There is another instance where a student appeals to a definition of linear pair to warrant a claim that the teacher insists follows from a theorem.

In the episode, the teacher raises the issue during the student's verbal recount of the proof. The student completes the recount up to step 5, at which point the teacher says: "OK. Please stop for a minute. What's the definition of linear pair just say [*sic*]?" The student responds by saying, "Wait, what?" The teacher continues: "What is a linear pair? It's two angles that do what?" After some additional prompting, the student says that a linear pair is two angles that share a ray and form a line. The teacher then says: "Technically [pause] he needs another little step in there. He needs to say angle 3 and angle 1 are a linear pair, and then he'd have definition of linear pair, and then he would need angle 2 and angle 4 are a linear pair, and he'd still have definition of linear pair." The teacher then states that what the student has for step 2 is actually a theorem that states that linear pairs form supplementary angles. To fix the proof, the teacher says that the student would need to add in another step and also edit the reason that warrants step 2. The statement that needs to be added is a statement that the angles are linear pairs, warranted by the definition of linear pair. The supplementary nature of the angles could then be warranted by the supplements theorem.

The previous episodes of proof checking show teachers drawing distinctions between definitions and theorems in two different classrooms. These instances suggest that drawing such distinctions in the written statements of a proof is a level of detail that teachers expect in order for proofs to be recognized as valid. The remaining episode examines an instance where a teacher indicates details that do not need to be included in a proof.

Checking proofs (IV): on details given in the diagram

The conversation from episode CM-101802-5P-S4 (described above) turns toward a consideration of what information is given by diagrams. The first step of the proof (see Figure 26, above) is the statement "△ABC". The reason that warrants this statement is: "Given". A student asks, "Is step 1 necessary?" The teacher says, "Um, well, when I wrote the problem, I had to have something up in the given, right? Uh [pause] so I would say, put it in there, alright?" The teacher elaborates on what she means: "Certainly you can use the angle sum theorem without first proving it's a triangle. That's one of those things you can assume from the drawing. If it looks like a triangle, it's a triangle."

A student follows up on this remark by asking: "So is it legal to do a whole proof without using given?" The teacher replies: "I suppose it is. Although remember you are given/you're given/in a way you're given a drawing in which you/you can make assumptions about things." The teacher then goes on to give an example of such a circumstance: She draws two intersecting lines on the board, labels their angles of intersection, and observes that, "without any given and without using the theorem", they would know how to prove that the vertical angles shown in the diagram are congruent.

The teacher's position that it is not necessary to prove that the figure that is represented in the diagram is a triangle suggests that there are limits to the kinds of details that need to be explicitly included in the written statements of a proof. In this case, the teacher does not require students to establish that the given figure is a triangle from more basic terms. Unlike the linear pair episodes considered above—in which teachers stated that students first needed to establish that the angles formed a linear pair before deducing that they were supplementary—in this case the teacher is saying that the triangle-ness of the figure is entailed by the diagram and does not require explicit proof.

All proofs ultimately rest on primitive notions[27]. It is therefore not the existence of an evidentiary threshold (Stylianides, 2007) that is significant, but rather how the line between necessary and unnecessary statements is being drawn when checking proofs in geometry. When the teacher says: "without any given", what she really means is "without any explicitly stated givens" or "with only diagrammatically stated givens." It is typical, for example, that separation, collinearity, existence, betweenness, and co-exact[28] properties (Manders, 2008) of figures are conveyed by diagrams when doing proofs in high school geometry (Herbst, Kosko, & Dimmel, 2013). This is a feature of the norms for reading diagrams (Weiss & Herbst, 2007) that help to reduce the complexity of proving tasks assigned to high school geometry students.

The fact that, when doing proofs in geometry, it is normative for proof problems to be stated in a diagrammatic register (Herbst et al., 2009) would suggest that teachers do not subject co-exact properties to scrutiny of the kind that was described above. The reason for the lack of scrutiny is not necessarily—or strictly—mathematical, since it would be possible for teachers to insist that the properties that are tacitly conveyed by the diagram be explicitly established. One method of uncovering the tacit properties that are given by a diagram of vertical angles (see Figure 30) is to state the problem using discipline-specific communication practices. So stated, the "given" for the vertical angles example would involve something along the lines of:

> Let *l* and *m* be any intersecting lines in the plane, and let *P* be their point of intersection. Let *A* and *C* be points on line *l,* and *B* and *D* be points on line *m*, such that *P* is between *A* and *C* and *P* is between *B* and *D*. Then $\angle APB \cong \angle CPD$ and $\angle APD \cong \angle BPC$.

---

[27] A primitive notion is an undefined term that is assumed to be true,

[28] A co-exact property of a diagram is a property that is unaffected by continuous variation of the diagram; that one region is included in another region, or that a region is bounded by a set of curves are examples of coexact properties (Manders, 2008, p. 92).

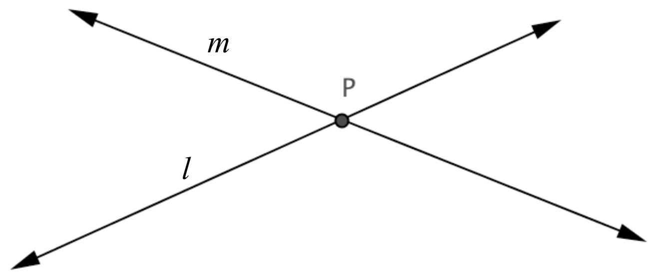

Figure 30: In the diagram, lines *l* and *m* intersect at point P, forming vertical angles.

That there exists a point of intersection for lines *l* and *m,* and that there exist points on the lines that satisfy the stipulated betweenness relations is some of the given information that is tacitly conveyed by the diagram, all of which depends on primitive notions of betweenness, incidence, and separation. It is not that there "isn't any given" for the vertical angles problem, but rather that the information that is given is not the kind of information that is usually (or expectedly) included in the written statements and reasons of a proof.

Summary of descriptions of *presenting* and *checking* proofs

The video episodes analysed above described the use of semiotic resources from different modes of communication to present and check proofs in geometry classrooms. In the case of presenting proofs, there were different levels of orchestration of semiotic resources from the range of available modes. Students were not expected to use discipline specific communication practices when presenting proofs to the class. In the case of checking the details of a proof, there were different evidentiary thresholds that were needed to establish when a proof was complete. The level of needed detail is mathematically specific and linked to different modes (i.e., writing versus diagramming). These comparable uses of semiotic resources in different episodes of geometry classrooms suggest that there could be normative ways of using semiotic resources when doing proofs in geometry. I call such normative ways of using semiotic resources *semiotic norms* and hypothesize semiotic norms for presenting and checking proofs in the next section.

Semiotic norms of presenting and checking proofs

In the episodes analyzed above, when a geometry student is *presenting a proof* to the class, the semiotic resources that have currency—that is, that may be cashed in the exchange of work for credit—are the written inscriptions that comprise the argument: i.e., the words, letters, and specialized symbols through which claims about a geometric figure are stated and warranted and the diagram to which these written inscriptions refer. Although a student could, in theory, use other semiotic resources such as speech, gestures, or other bodily movements to present the argument, what I observed in different episodes of geometry instruction is that the most important part of the presentation is the written proof. This suggests there are normative ways that students use the available channels of communication when presenting proofs to the class. Thus, one semiotic norm that I conjectured from the video episodes is the *channel norm*: when presenting proofs in geometry, students are expected to present a proof that is entirely contained in written statements and reasons.

The video episodes described above show that student proof-presenters appeared to be operating under the tacit requirement that *cashing* (i.e., exchanging for credit) a proof called for producing a complete written record of a proof. In instances where students used other modes of communication to recount a proof that was presented on the board—such as in Megan's classroom—the information that students communicated through speaking and gesturing was redundant with what was communicated by the written statements and reasons of the proof. The hypothesis that such a central emphasis on the written statements and reasons of the proof is normative for student proof presenters is warranted by the fact that the teachers did not mark the students' reliance on the written mode of communication as a behavior that needed to be repaired. That is, for the teachers and students in the videos, the reliance on the written mode of communication was routine, usual—i.e., *nothing out of the ordinary.*

The channel norm concerns the mode of the semiotic currency that is valid when students present proofs in geometry. An additional semiotic norm concerns the logical versus temporal sequencing of the written statements and reasons of the proof that a student presents. The descriptions of students presenting proofs in geometry classrooms in the video episodes analyzed above suggest that, when a student is presenting a proof, the student acts as a transcriber. What this means, in practice, is that the argument a student creates at the public canvas is not required to be logically reasonable until the transcription is completed. The order in which a student transcriber recreates the different parts of the proof need not match the sequence that would be required in order for the semiotic resources used in the argument to be intelligible. For example, a student might write the statements and reasons of a proof out in their entirety *before* drawing or labeling the diagram that is needed for those statements to be meaningful[29]. The fact that students normatively act as transcribers when presenting proofs and that a transcribed proof is not reviewable until the transcription is completed is captured by what I call the *sequence norm*: when students are presenting a proof, they may present the semiotic resources that comprise the proof in any order they choose; and usually, the order that students use is that which most easily facilitates the literal transcription of semiotic resources.

The channel norm and sequence norm are semiotic norms that pertain to how proofs are presented in geometry classrooms. One reason to present a proof to the class is to review the proof and check its written details to ensure that the proof may be accepted (by the teacher) as a

[29] This happens in the first video of students presenting proofs that was described above.

valid mathematical argument. In the language of instructional exchanges, *proof checking* is the act of reviewing the written details of a proof to determine whether it may be cashed for a claim on the knowledge item that is at stake in the proof. During this proof checking activity, the teacher and students in the class take turns scrutinizing the written steps of a proof that has been presented to the class.

In the video episodes of proof checking in different geometry classrooms described above, there were recurring instances of details of the proof being insufficient under such scrutiny. These included instances when conceptual entailments—such as the conclusion that two angles that form a linear pair are supplementary—were not unpacked into more basic steps (i.e., a statement that identifies such angles as being a linear pair followed by a statement that angles forming a linear pair are supplementary, by the supplements theorem) and instances when distinctions between geometric objects and their measures (such as a segment versus the length of a segment) were not strictly enforced. Since the kinds of details that were scrutinized recurred in different geometry classrooms, it is reasonable to hypothesize that there are normative ways of checking the details of a proof in geometry—i.e., a *details norm.*

The *details* norm is the hypothesis that teachers check the details of a proof in mathematically specific ways that are linked to different semiotic modes. The video episodes of proof-checking suggest that teachers check to ensure the inclusion of two kinds of details in the written statements and reasons of a proof: (1) that conceptual entailments are unpacked into more basic steps—e.g., stating that the angles formed by an angle bisector are congruent, by the definition of angle bisector; and (2) that objects are distinguished from their properties—e.g., stating that two segments have the same measure, by the definition of congruence. Such routines for checking proofs attend to the use of written statements (e.g., the definition of midpoint) to warrant other written statements (the congruence of two segments).

Although written statements that could be unpacked into more basic written statements are an area of emphasis when the details of a proof were checked, how diagrams were specified was not subject to scrutiny when proofs were checked. That is, whether a point was interior or exterior to a figure, or whether there exists a point of intersection for two lines or line segments was not a detail that students included or that teachers required in the reviewed episodes of doing proofs. In the videos, teachers routinely made comments to the effect of "no detail is too small", yet the domain of this edict appears to be limited to the details of the written statements in the proof. The theoretical details—such as: an appeal to axioms of completeness, continuity, or betweenness—that might warrant a particular realization of a geometric figure in a diagram are not required to be included in a proof in order for it to be cashed.

Chapter Summary

This chapter drew on video records of geometry classrooms to describe how proofs are presented and checked in geometry classrooms. In the sample of videos in the lesson corpus, teachers have comparable expectations for how students use modes of communication to represent their work when dong proofs. The small sample limits the range of the conclusions that are warranted from the descriptions of the video episodes reported above. The most that could be said is that, across the episodes described above, there appear to be regular ways that modes of communication are used when proofs are presented and checked in geometry classrooms.

From the descriptions of video episodes, I hypothesized three regularities of how modes of communication are used when presenting and checking proofs that I called the channel, sequence, and details norms.. The acts of communication by teachers and students that underlie these hypothetical norms were observed in different episodes of three geometry classrooms taught by different teachers. The force of the hypotheses rests not on the number of actors who were observed acting in accordance with the norms, but rather on the fact that teachers and students in different classrooms showed evidence of acting in similar ways. It is not hard to imagine that each classroom could have developed different routines for presenting and checking proofs. That, across the classrooms, students behaved as transcribers when presenting proofs and teachers keyed in on the same types of missing details in the written arguments of proofs suggests that these are recurring patterns of activity.

Aside from their status as hypotheses derived from the analysis of video episodes of geometry classrooms, the hypothesized semiotic norms are also reasonable expectations for teachers to have. All three of the hypothesized norms are linked to the written mode of communication. The discipline of mathematics has an adaptive, economical written symbol system of representation that continually grows to incorporate new mathematical knowledge (Thurston, 1994; Tappenden, 2008; Sfard, 2008). Writing is indispensable to the work that mathematicians do (Greiffenhagen, 2014). That the written mode of communication would draw attention when proofs are presented and checked in geometry classrooms could thus be seen as being consistent with mathematics teachers fulfilling their obligation to the discipline of mathematics (Herbst & Chazan, 2012). The *channel, sequence,* and *details* norms could each be understood as regularities that help teachers ensure that the development of students' proof-writing skills are honed during the course of learning to do proofs in geometry.

That explicitly helping students develop other modes of communication would receive less attention in geometry classrooms is also reasonable, given the realities of limited time and standardized testing that teachers need to contend with (Nolan, 2012). For teachers to provide students with explicit guidance about how they are using the verbal and gestural modes of communication would require designing tasks that would target these skills. In some ways, that students come to the board to present proofs indicates that there could be opportunities for developing such tasks in geometry classrooms. But if students were expected to develop the mathematical communication skills that approximate the fluid coordination of modes that are used during *chalk talk,* it would require more investment of time on the part of the teacher into activities during which students could rehearse—and get real-time constructive comments on—these skills. I am not suggesting that it is infeasible for mathematics teachers to find ways to incorporate such opportunities into their classrooms. Rather, I am pointing out that, given the challenges of implementing tasks that would incorporate these other skills in a central—rather than ancillary—way, the emphasis on presenting complete, written proofs in geometry classrooms is reasonable.

The role that diagrammatic modes of communication play in the *details* norm is somewhat curious, given (1) the notorious legacy of flawed diagrams in the history of geometry (Barwise & Etchemendy, 1996) and (2) the disciplinary practice of skipping "obvious" steps in a proof (Davis, 1972). The *details* norm seems to turn these points on their head: Teachers in the episodes were perfectly willing to assume that particular obvious relationships were given directly by diagrams yet were reluctant to accept proofs that elided obvious written steps. I revisit this issue during the reporting of the experimental part of the study. The next chapter describes a method for experimentally investigating the extent to which secondary teachers recognize the regularities that were hypothesized in this chapter as norms.

CHAPTER 4:

A METHOD FOR EXPERIMENTALLY TESTING RECOGNITION OF HYPOTHESIZED CLASSROOM NORMS

I describe here a method to experimentally test the extent to which teachers recognize hypothesized classroom norms. The teachers that are the subject of this study are secondary mathematics teachers. The norms that are the target of the study are examples of what I call *semiotic norms*: normative ways that semiotic resources are used within instructional situations.. The semiotic norms that were investigated for this part of the study are hypothesized to be norms of the instructional situation of *doing proofs in geometry*, episodes of which were described in Chapter 3. In what follows, I describe an experimental variant of the classic breaching experiment (Garfinkel, 1963, 1970; Mehan & Wood, 1975; Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009) technique that I used to empirically test for secondary mathematics teachers' recognition of semiotic norms.

The challenge of seeing what usually is

The method I used for experimentally testing the extent to which teachers recognize classroom norms is a refinement of methods that researchers have used to empirically study the tacit and generally invisible routines—also called *background expectancies* by Garfinkel (1963) and *situational norms* by Herbst (2006)—through which social order in classrooms is produced. The invisibility of routines has been a point of emphasis for the ethnomethodological account of social reality (Garfinkel, 1963; Mehan & Wood, 1975; Atkinson 1987) and identified as a challenge by scholars who aim to describe the social organization of classrooms *as they appear* *to those who participate in the social construction of classrooms*—e.g., teachers, students (Mehan, 1979; Herbst, 2006; Herbst & Chazan, 2003; Herbst & Chazan, 2011; Aaron & Herbst, 2012). Herbst, Nachlieli, and Chazan (2011) summarize the challenge: “Inasmuch as some of the elements that help explain teaching action are tacit, it is important for the field to have techniques that elicit them without a mediating process in which they become explicit (which would be the case using survey instruments)” (p. 223).

Following Garfinkel (1963) and Herbst and Chazan (2003), I use *norm* to identify an aspect of a social situation that not only regularly happens, but that those participating in that situation *expect* to happen. Furthermore, such expectancies of a situation are generally transparent to those participating in the situation—transparent in the ethnomethodological sense of being the unremarkable constituents of the shared everyday activities that comprise the situation (Herbst, Nachlieli, & Chazan, 2011). Some examples will help to illustrate the transparency of such routines and highlight the methodological challenges of investigating them.

When walking outside in the cold, the fact that other pedestrians are bundled up is unremarkable; bundling up is just something people do (and expect others to do) when bracing the cold. If, on a winter morning, I were asked by an investigator to give an account of my walk into work, it likely would not occur to me to describe the winter clothing that I was wearing, or to comment on the winter clothing worn by other pedestrians I passed on the way. The shared expectation that people wear warm clothing when it is cold outside is so commonplace as to be transparent and totally unremarkable.

Another example that illustrates the transparency of what is expected comes from the world of competitive sports. During broadcasts of professional sporting events—such as American football games—there is a “play-by-play” announcer who describes what is happening in the game as it happens. While this announcer provides a running record of what is taking place during the game as the game is played, there are many details about the scene that are left out of the play-by-play account. For example, the announcer does not alert the listener that each team is wearing their appropriate uniforms, or that the correct number of players is on the field for each team during each play, or that the game is being played with a regulation ball. These aspects of the situation—while mutually expected by all parties directly competing in the contest, as well as by those who are listening to a broadcast of the game—are part of the *background expectancies* (Garfinkel, 1963) of the game being played.

These examples help to indicate the methodological challenges inherent to providing an account of some aspects that those who routinely participate in an activity expect will happen during the activity. Framing the expectancies as survey questions essentially telegraphs the targeted expectancy to the person taking the survey. Recovering what people expect in situations by asking them to describe the situation is insufficient, because the things they most expect are also quite possibly the things they are least likely to describe. To take an extreme example: if one were preparing a meeting room, one would never write “air to breath” on a to-do list of things needed for the room in order to host the meeting. The availability of breathable air is such a baseline expectation no one would consider this a requirement for the meeting. It would be absurd to conclude that a meeting organizer overlooked the necessity of breathable air just because it wasn’t explicitly written on a list.

As an alternative to querying participants directly, through observation one could develop hypotheses about the things that regularly happen in a situation. This method is at the heart of the study of *teaching scripts* (Stigler & Hiebert, 1999; Jacobs & Morita, 2002; Seidel & Prenzel, 2006): stable patterns of classroom activity that teachers recognize as typical instances of particular kinds of instruction (e.g., mathematics lessons, physics lessons). But once these hypotheses are generated, the investigator would be left with the challenge of determining whether the observed regularity is, in fact, expected to happen by those who participate in the situation. To take a trivial example: it may be that one of my coworkers always writes notes on college-ruled paper. If someone were observing our weekly meetings to the end of noting regularities, this fact is something that an outsider might observe. Yet despite its regularity in the situation, the quality of paper my coworker uses for note-taking is not likely to be something that I expect of the situation—my coworker could one day use a different kind of paper, and there is a fair chance that this change would not affect in any way the unfolding of the meeting. This underscores the point that not all things that regularly happen are expected to happen.

But what if something that participants in a situation generally expect to happen fails to happen? How might participants in the situation respond? Conducting this thought experiment led Garfinkel (1963) to propose the "breaching experiment technique" as a means for making the routine, expected, and tacit social order an object of sociological study. The behavior of a computer science professor during the "peanut butter sandwich" task (Davis & Rebelsky, 2007)—in which students are to provide an explicit set of instructions for assembling a peanut butter sandwich—is effectively a breaching demonstration, whereby the professor deliberately ignores the tacit, unspoken meanings of the instructions the students provide—for instance: by taking "open the bread" as an invitation to tear a piece of bread apart from its ends (Davis & Rebelsky, 2007)—to the end of illustrating the ambiguity inherent in those instructions.

Reconsidering the example of walking to work on a winter morning provides another illustration of the technique. Imagine being outside on a winter day and coming upon a person who was wearing shorts, short sleeves, and sandals. A person so attired would be in violation of the tacit expectations that people have of themselves and others when it is cold outside. While there are no laws prohibiting such attire, a person so dressed would likely stand out during one's otherwise routine walk to work, e.g.:

[reporting to a coworker] I saw a man wearing shorts and sandals on my walk in!

[coworker responds] In this weather? What was he thinking?

At this juncture in the conversation, the person making the report might offer an attempt to explain or otherwise normalize the odd behavior, e.g., "Maybe he was locked out of his house. I should have checked to make sure he was ok". Garfinkel (1963) described such efforts to restore normality as *repairs* to the situation. This use of "repair" does not mean "fix" as much as it does "reclassify", i.e., *in what kind of situation would such [otherwise unexpected] behavior be routine?* That violations of the generally unseen social order evoke repairs from social actors that can reveal aspects of that order is the heart of the "breaching experiment" technique (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009; Herbst, Nachlieli, & Chazan, 2011).

Garfinkel called planned violations of the expected social order "breaching experiments", though by Garfinkel's (1963) own account, the use of the term "experiment" is a misnomer: "Despite their procedural emphasis, my studies are not properly speaking experimental. They are demonstrations, designed…as 'aids to a sluggish imagination.' I have found that they produce reflections through which the *strangeness* of an *obstinately familiar* world can be detected." (emphasis added, p. 227). The breaching experiment technique provides a means for probing what participants in social situations expect of the situation and those others that are participating in it.

The breaching experiments described by Garfinkel (1963) and others (e.g., Rafalovich, 2006) were performed by people—as the breaching agent—on people—as unwitting participants—and tended to result in at least minor psychological injury to all parties involved (Gregory, 1982). For example, Garfinkel (1963) describes an argument that erupted between a student and her husband when the student breached one of their conversation norms: the student repeatedly asked her husband “what do you mean?” while he was going through his typical narration of the events of the day. In another case, Garfinkel (1963) recounts the cool reception a student received from members of her immediate family, after she had behaved as if she were a boarder in her own home—thus breaching some of the norms of what it means to be someone’s child (see Garfinkel, 1963, for other examples).

In these and other cases, those who experienced the breaching behavior “…vigorously sought to make the strange actions intelligible and to restore the situation to normal appearances." (Garfinkel, 1963, p. 232). The repair strategies described by Garfinkel (1963) tended to involve anger, bewilderment, or anxiety; subjects reacted quickly and strongly to breaching behavior. Garfinkel (1963) conjectured that such forceful responses were a defense of the “natural facts of life” or “routine social order” that was immediately threatened by those perpetrating the breach (p. 236). On account of these potentially strong reactions on the part of those who experienced the breaches, Mehan and Wood (1975) warned that: “Interested persons are advised not to undertake any new breaching studies. It is immoral to inflict them on others” (p. 113). Of such potentially strong reactions, Garfinkel (1963) posited that “…the firmer a societal member's grasp of [*sic*] What Anyone Like Us Necessarily Knows, the more severe should be his disturbance when ‘natural facts of life’ are impugned for him as a depiction of his real circumstances.” (p. 236) Mehan and Wood’s (1975) warning notwithstanding, the technique of breaching the tacit, but expected, routines through which social order is produced remains an effective method for gauging the extent to which participants in a given social activity expect such routines to be followed. As a technique for studying the social order of classrooms, breaching experiments offer the researcher a window into how classroom order is produced from the perspective of those who actively produce that order (Herbst & Chazan, 2003). On this point, I turn to a review of how breaching experiments have been adapted for use to study the social order of classrooms.

Breaching experiments in education research

The breaching experiment technique described by Garfinkel to make the tacit social order witnessable has the advantage of simplicity—e.g., no lab equipment required, anyone can perform a breaching act and witness the reactions of those who experience the breach—but the disadvantages of (1) inflicting psychological injury on those who propagate and experience the breach, and (2) not having experimental controls. Education researchers (Herbst & Chazan, 2003; Herbst & Chazan, 2006; Herbst, Nachlieli, & Chazan, 2011; Herbst, Aaron, Dimmel, & Erickson, 2013) modified the classic breaching experiment technique in two ways to adapt it for studying the tacit social order of classrooms.

To minimize the risk of inflicting psychological injury, researchers (e.g., Herbst 2003; Herbst, 2006) designed breaches that departed from the expected classroom order but that were still justifiable on disciplinary or pedagogical grounds. For example, working in collaboration with high school geometry teachers, Herbst (2003, 2006) designed a series of replacement lessons in which the teacher departs from hypothesized normative ways that teachers engage students in doing proofs in geometry by moves that straddle the boundary "between the customary and the unusual" (Nachlieli, Herbst, & González, 2009, p. 432). In such episodes, even though the teacher departs from what students and teachers of high school geometry expect to happen when doing proofs in geometry, the breaches were not occasions for harming the students in the classroom. Rather, the breaches created opportunities for students to engage in different kinds of mathematical work; e.g., coming up with a statement that is to be proved (when this statement would typically be provided by the teacher) (Herbst & Chazan 2003). While asking a teacher to solicit different kinds of mathematical work—e.g., making a statement about how the angle bisectors of a quadrilateral relate to each other versus proving some particular relationship about those angle bisectors (Herbst & Chazan, 2003)—from students is a breach of what is expected in a geometry classroom, the potential for psychological harm as a consequence of this breach is virtually nil. Or at least, the potential for injury with such replacement lessons is not demonstrably greater than the potential for injury that comes with any mathematics lesson.

To increase experimental control, researchers (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009; Herbst, Nachlieli, & Chazan, 2011; Herbst, Aaron, Dimmel & Erickson, 2013) have devised the technique of a *virtual* breaching experiment. In such experiments, social actors who are familiar with particular types of situations are shown a representation of a situation in which a norm of that situation is breached. In one of the earliest examples of such virtual breaching experiments, high school geometry teachers were shown a video episode of an actual mathematics classroom in which the teacher in the episode departs from the normative ways of doing proofs in geometry (Herbst & Chazan, 2003). In such a design, different groups of teachers could be shown the same breaching stimulus, and their responses to the episode (i.e., their conversations that followed) could be video and audio recorded, and analyzed later.

The use of records of classroom episodes as probes for discussions among groups of teachers exemplifies the potential of using representations of practice cyclically, to generate testable hypotheses from qualitative data. Jacobs, Kawanaka, and Stigler (1999) outline a method for using video records to generate such data:

> First continual viewings of the tapes help generate informed ideas and analyses. Next, quantitative analysis allows for the validation of discoveries made by watching the videos. Then, qualitative analysis leads to clearer interpretations of the results from the statistical analyses. Looping through the cycle many times helps to generate new questions, refine coding systems, and locate footage that can serve to exemplify particular findings. (p. 719)

Pushing this technique further, Herbst, Nachlieli, and Chazan (2011) argue that using representations of instruction that use a simplified graphics language to depict classroom scenes offers researchers even more affordances for probing how teachers see classroom activity: "To the extent that the cartoon-representation of people and settings can be controlled to abstract some elements of context and focus on chosen elements of practice, our technique appears to be more malleable than video for the development of research instruments" (p. 223). For the purposes of designing virtual breaching experiments, the use of animations has the advantage that disruptions in classroom routines can be captured without the infelicities of teachers enacting and students experiencing the breach (Herbst, Nachlieli, & Chazan, 2011).

To demonstrate the viability of this approach, Herbst, Nachlieli, and Chazan (2011) developed animations of episodes of mathematics instruction to use as probes for teacher discussion groups. These representations were scripted to depict the teacher in the classroom breaching a hypothesized norm[30]. The episodes were shown to groups of teachers and their conversations around the episode were video (and audio) recorded. Participants' reactions to the episode were analyzed for evidence that participants noticed the breach of the norm and for the

[30] For this study, the target norm is that, when installing a new theorem in geometry, a teacher needs to explicitly sanction that what has been proved is a theorem that can be used as such in future proofs.

strategies that participants offered for repairing (i.e., normalizing) the situation. The quantifiable data generated from this qualitative analysis was a series of ratios of the rates at which target repair strategies occur in intervals of conversation (see Herbst, Nachlieli, & Chazan, 2011, p. 235 for details).

Herbst, Aaron, Dimmel, and Erickson (2013) further developed the method of a virtual breaching experiment through multimedia questionnaires. In their design, participants viewed an instructional episode then answered a series of open-and closed-ended questions via an online research environment[31]. Different episodes of geometry instruction were scripted to show different ways of breaching situational norms. The instructional episodes were realized through storyboards that used cartoon characters to represent teachers and students (Herbst, Chazan, Chen, Chieu, & Weiss, 2011). The study provided empirical evidence that such representations of classroom situations could be used as a probe to gauge teachers' recognition of hypothesized norms.

In assessing the limitations of their study, Herbst et al. (2013) concluded that it was asking too much of participants to respond with repairs that would specifically point to the norm that had been breached. Instead, consistent with previous accounts of participants' reactions when expected routines were breached (Garfinkel, 1963), it was more likely that participants might, in a manner of speaking, "throw a fit[32]", and respond in a general negative way to the fact that some tacit expectation they had was not being met. In theory, such general negative reactions would be evident in responses to closed-ended rating questions. But since the episodes

[31] This is the Lesson*Sketch* environment (www.lessonsketch.org). Lesson*Sketch* was also used to gather the data for this study.
[32] P. Herbst, internal communication.

showed not only the breach of the target norm but also other instructional actions[33], there is the chance that participants were responding negatively to some other aspect of the storyboard. The experimental design I describe in the next section attempts to control for these potentially confounding effects on participants' ratings.

A Virtual Breaching Experiment with Control

The technique for investigating the extent to which teachers recognize a hypothesized classroom norm combines a virtual breaching experiment (Herbst & Chazan, 2003; Nachlieli, Herbst, & González, 2009)—described above—with planned, randomized comparisons (Shadish, Cook, & Campbell, 2002). These comparisons were planned both within groups (that is, participants in the same experimental condition had their responses to different types of episodes compared) and between groups (that is, participants were randomly assigned to experimental conditions, and within each condition, participants viewed different episodes of instruction). The use of customizable cartoon graphics to create representations of episodes of doing proofs in geometry raised the possibility of introducing control conditions into the virtual breaching experiment (Herbst, Chazan, Chen, Chieu, & Weiss, 2011). The next section elaborates on how such controls were realized in the experimental design.

Comparisons with control: principles for scripting storyboards

The study described here uses virtual breaching experiments with controls to gauge the extent to which high school mathematics teachers notice breaches of semiotic norms in episodes of geometry instruction. The semiotic norms that were targeted for this part of the study are the details norm (DN) and the sequence norm (SN) that were described in Chapter 3. The experimental part of the study focuses on the details and sequence norm because I believed

[33] The inclusion of other such instructional actions were necessary to make the storyboard viable as an episode of instruction.

reactions to breaches of these norms would be more varied than reactions to breaches of the channel norm. That is, using the written channel is such a bedrock feature of the common sense world of doing proofs that if participants were shown an episode of instruction in which students communicate a proof using primarily non-written channels—such as speaking or gesturing—their reactions would very likely be strongly negative. That the written channel plays a central role in the communication of proof is not surprising; however testing whether teachers recognize that different kinds of written details are more crucial than others when checking a proof (the details norm) or that a proof is not subject to scrutiny until it is complete (sequence norm) provide opportunities to gain a more nuanced understanding of how teachers and students exchange work for knowledge claims when doing proofs in geometry.

Below, I summarize the storyboards I developed to gauge the extent to which teachers recognize the details and sequence norms. But before diving into the specific descriptions of each storyboard, I want to highlight three general principles that framed the design of the breaching storyboards. One is what I will call the principle of *reasonable departure*: The ways in which a teacher breaches a target norm should not be egregious, arbitrary breaks with what is expected in the classroom, but rather should be reasonable deviations from what ordinarily happens. This principle—while not deliberately applied during the breaching demonstrations of the 1960s—is at play in the design of previous virtual breaching experiments (Herbst & Chazan, 2003; Herbst, Nachlieli, & Chazan, 2011; Herbst, Aaron, Dimmel, & Erickson). Adhering to this principle is a means of ensuring that participants recognize the episode as a viable instance of mathematics instruction.

The principle of reasonable departure is important because the virtual breaching experiment with control instrument—described in more detail below—uses responses to closed-ended rating questions as evidence that participants are reacting to breaches of what they expect will happen in particular instructional situations. That evidence is only compelling if the teacher's departure from the norm is justifiable[34]. If a departure is too extreme—for instance, a teacher breaches a norm by stating the givens of a proof as a limerick, or responds to a student by throwing a marker—the fact that there are negative responses would not be surprising. But if participants notice and react to a teacher taking an alternative, but still justifiable, tack, then this is an indication that they are responding to the absence of something they expect should happen in the depicted situation. Below, I describe how each of the breaching storyboards I scripted satisfies the *reasonable departure* requirement.

The second general principle is that there should be routine teaching actions unrelated to the target norm common to every breach/control pair. I will call this the principle of *common ground.* Including common ground in breach/control pairs serves two purposes. One, it helps ensure that there is not too much attention being called to the moments in a storyboard where the teacher breaches the norm. Besides breaching the norm, the teacher makes some other, routine move—e.g., explains a concept, answers a question—that could draw the attention of a person viewing the storyboard. Two, including common ground across every breach/control pair introduces the possibility of asking rating questions that target specific moments in the storyboard. The instrumentation of such questions is described below, but the general idea is that the availability of common ground means that, in addition to a general rating question, participants could rate the actions of the teacher at specific places in the storyboard—e.g., a place where the teacher breaches (or does not breach) the norm and a place where the teacher does

[34] Herbst, Aaron, Dimmel, & Erickson (2013) designed breaching storyboards in which the departures were hypothesized to be justifiable by appeal to the professional obligations of mathematics teaching (Herbst & Chazan, 2011).

some other action that is unrelated to the norm and common to each storyboard. The responses to these targeted rating questions could then be compared across conditions, with the common ground functioning as a control.

A third general principle that underlies the design of the storyboards is the principle of *minimal variation*. The use of storyboards with cartoon graphics offered the affordance that any episode of instruction could have multiple, minimally different realizations as a storyboard. Thus, every episode of instruction that I scripted was designed to have two instantiations: a breach version and a control version. The breach and control instantiations of an instructional episode were identical except during those places in the storyboard where the teacher complies with or departs from the norm. The aim of creating storyboards across breach and control conditions that were minimally different from each other was to be able to link how participants rated the work of the teacher in a storyboard to whether the teacher was shown as breaching or complying with the target norm.

For example, if storyboard A depicts a teacher breaching a norm, then storyboard A' would be its *control dual*—an image sequence that is identical to storyboard A except for the images in the sequence that depict the teacher breaching the norm. One group of participants would view and respond to storyboard A while a different group of participants would view and respond to storyboard A'. The reactions to storyboard A could be compared to the reactions to storyboard A', and since everything in these storyboards was the same except the part of the storyboard where the teacher breaches (or does not breach) a target norm, the minimally different design of the storyboards would provide a warrant for concluding that participants were responding to the breach, as opposed to some other aspect of the storyboard. Additional controls (see below) were built into the instrument design to strengthen the claim that participants' responses to closed-ended questions were most strongly influenced by viewing condition (i.e., breach or control) as opposed to some other aspect of the storyboard.

Scripting storyboards: creating breach/control pairs

The episodes were scripted image sequences that were adaptations of geometry lessons[35] that were based on video recordings of classrooms doing proofs. By adaptation, I mean that I identified essential exchanges between students and teachers in actual classroom videos (analyzed in Chapter 3) and then scripted a viable storyboard of geometry instruction in which the target exchange could naturally occur. The specific influences for each of the different storyboards are described below. I raise the issue here to call attention to this way of designing research probes as an instance of the cyclical use of records of practice (Jacobs, Kawanaka, & Stigler, 1999, described above).

As a planned comparison study, participants were randomly assigned to treatment or control conditions in which the teacher in the storyboard episode breaches (treatment) or complies with (control) a target semiotic norm. The breaching storyboards targeted either the details norm or the sequence norm (not both), while the compliant storyboards did not breach any classroom norms. Rather, the control storyboards depicted what routinely happens when presenting or checking proofs in geometry.

A total of 10 storyboards were scripted: 4 that target the details norm and 6 that target the sequence norm. Each of these 10 storyboards had a breach instantiation and a control instantiation. Thus there were a total of 20 instructional episodes that were represented using storyboards. The storyboards were inspired by classroom video. By basing the storyboards on actual episodes of instruction, the goal was to create probes that would be recognizable to

[35] From the lesson corpus that was described in Chapter 3.

participants as viable episodes of geometry classrooms. In the next sections, I describe the sets of storyboards that were created to test each of the target norms.

Developing storyboards that breach the details norm

I developed 4 storyboards that could be used to probe the extent to which teachers recognize the details norm. In classroom videos of geometry classes checking proofs, students and teachers scrutinized similar kinds of details of the written record of a proof. The warrant for the scrutiny is that omissions are tantamount to gaps in a proof. While certain kinds of details seemed to always be the subject of scrutiny (more on these below), other kinds of details seemed always to be not acknowledged. The details that went unacknowledged were the specification of geometric properties that are normatively conveyed by the diagram, such as properties of incidence, separation, or order (Herbst, Kosko, & Dimmel, 2013). Problematizing the existence of a point of intersection between two lines, for example, is not within the domain of what is usually scrutinized in a proof. Insisting that the existence of such a point of intersection be explicitly warranted by an appeal to a proposition—despite the availability such a proposition in standard developments of Euclidean geometry—appears to be outside the scope of the routine work of checking proofs in geometry.

That some details are scrutinized—on the grounds that there can be no gaps in a proof—while others are overlooked suggests that there are two ways to script episodes of instruction to investigate the hypothesis that there are normative ways of checking proofs in geometry classrooms. One strategy is to script storyboards in which a teacher breaches the details norm by accepting—i.e., cashing—a proof that overlooks a detail that is usually subject to scrutiny. The second strategy is to script storyboards in which the teacher breaches the details norm by asking for written details that are normatively telegraphed by the diagram.

In the first case, the teacher could be seen as breaching the norm by accepting a proof that has *less detail* than what is normatively required. In the second case, the teacher could be seen as breaching the norm by insisting on *more detail* than what is normatively required. It was important to develop storyboards that could be used to test both hypotheses to be able to argue that the norm is not actually a generic one (insisting on detail, no matter what detail), but rather a mathematically specific one.

I developed 2 storyboards in which the teacher breaches the norm by accepting proofs with less detail and 2 storyboards in which the teacher breaches the norm by insisting that proofs contain more details. Below, I summarize the breach and control instantiations of each storyboard. I also recount the events that occurred in actual geometry classrooms that inspired the storyboards.

Developing storyboards that breach the norm by accepting less detail

The 2 storyboards that target the *less details* way of breaching the details norm were inspired by episodes of geometry instruction in which a teacher argues for the need to include explicit details in the written argument of a proof. The first storyboard is based on an exchange between a teacher and student about whether it is necessary to include a written statement that deduces the congruence of two segments from the definition of midpoint. This exchange occurs in episode LV-111303-4P-S4 and was described already in Chapter 3.

The episode suggests several possibilities for ways to script a storyboard that could target the norm about the level of detail that teachers expect when checking proofs. In the storyboard, a student asks the teacher about whether it would be ok to skip some of the steps in the proof. One interpretation of what the student professes to skip is that the student writes what is given, then skips directly to the statement that the required triangles are congruent by ASA. From the student’s perspective, such a move could be seen as reasonable because the given statements entail the necessary congruence relations, and by selecting ASA as the appropriate congruence criterion to apply, the student is communicating that the student knows what parts of the triangles will be used to establish their congruence. The fact that the teacher explains the necessity of all three intermediate steps in the proof, together with the stroke the teacher adds to the argument that visually bypasses those intermediate steps, provides some evidence that this is what the student may have confessed to. One way of breaching the norm would then be to show the teacher accepting a proof in which the student only writes the given and the conclusion.

An alternative interpretation is that the student skipped the step that establishes the congruence of JL and LA from the definition of midpoint. There is evidence for this because the teacher states “then you’ve left out a big step”—as opposed to steps—and the teacher begins her explanation of what is necessary with the statement that the segments are congruent, warranted by the definition of midpoint. A second way of breaching the norm would thus be to show a teacher accepting an argument in which the student does not explicitly establish the congruence of the segments from the definition of midpoint.

A final way of breaching the norm is suggested by the comments that the teacher makes about accepting “somewhat abbreviated reasons.” The teacher states that she will accept “corr. <’s” as a sufficient warrant to establish the congruence of the corresponding angles, but warns that other teachers might want reasons that are more specific. If it is the case that other teachers would insist on more specific reasons—for instance, insist that the student state the corresponding angles theorem in its entirety—then the episode, as is, shows a third way of breaching the norm.

The first possibility—a student writes only the given and an appropriately warranted conclusion—is not likely to be seen as a *reasonable departure* from what teachers in geometry classrooms normally do. While it may not be out of the ordinary for a student to make that kind of argument, it seems highly unusual that a teacher would recognize such an argument as an instance of doing a proof. Were a student to make such an argument, even the student would likely regard such a maneuver as a joke or an exception to what is typically expected. The third possibility presents the opposite challenge: since the teacher in the episode herself accepted the abbreviated reasons, it seems likely that making such small allowances is in line with the norms for checking the details of a proof in geometry classrooms.[36] The first possibility is likely too extreme; the third possibility is likely too subtle. This leaves the second possibility—in which the teacher accepts a proof wherein the student does not explicitly establish the congruence of the midpoint segments—as the inspiration for storyboard 26001, the first storyboard that targets the less details norm.

The second storyboard that targets the less details norm—storyboard 26002—is a variation on the theme described above. The inspiration comes from the instances that were described in Chapter 3 during which the teacher strictly adheres to the distinction between angles and measures of angles. That teachers insist on the distinction between angles and measures of angles suggests that a teacher could breach the details norm by cashing a proof that that doesn't heed this distinction.

Synopsis of storyboards 26001 and 26011. Figure 31 shows the problem that was depicted to target the less details norm in storyboard 26001.

---

[36] This observation will be reconsidered in the reporting of results and discussion.

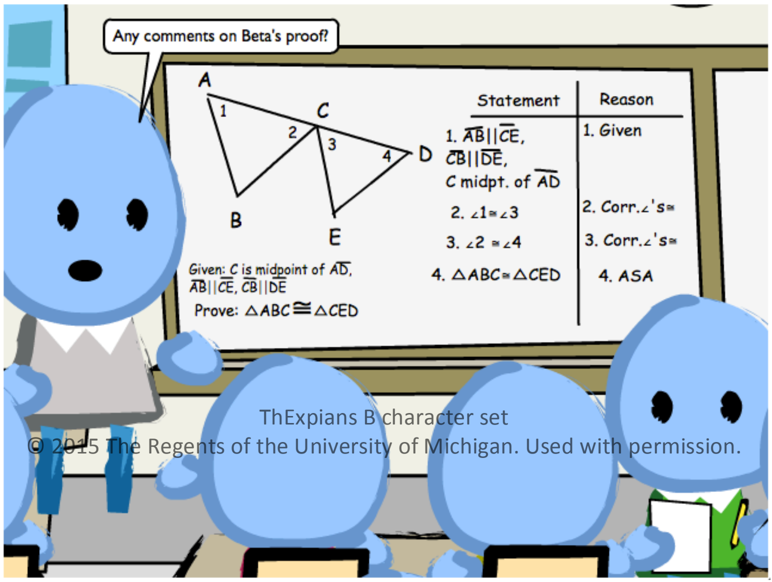

Figure 31: Problem used to target the *details* norm in storyboard 26001.

In storyboard 26001, Beta (a student) writes a proof on the board. The teacher asks for comments, and a different student asks: "Why does it say ASA? All it shows is two congruent angles." To breach the norm, the teacher replies: "Well…C is the midpoint of AD. So we know what Beta means. Other comments?" The control dual[37] of this storyboard—storyboard 26011—shows the same proof, with a student raising the same objection, only in this storyboard, the teacher recognizes the omission and corrects the proof. The teacher says: "We need a separate statement to show that AC and CD are congruent. " The teacher then corrects the proof by adding this missing step. In both versions of the storyboard, the episode continues and the teacher asks if there are any other comments on the proof. The common ground across storyboards 26001 (breach) and 26011 (control) is a student who says "I did it with AAS" and the consequent discussion.

Synopsis of storyboards 26002 and 26012. Figure 32 shows the problem that was depicted to target the *less details* norm in storyboard 26002.

[37] The *control dual* of a storyboard is defined above. It is the minimally different, alternative version of a storyboard in which the teacher complies with, rather than breaches, the norm.

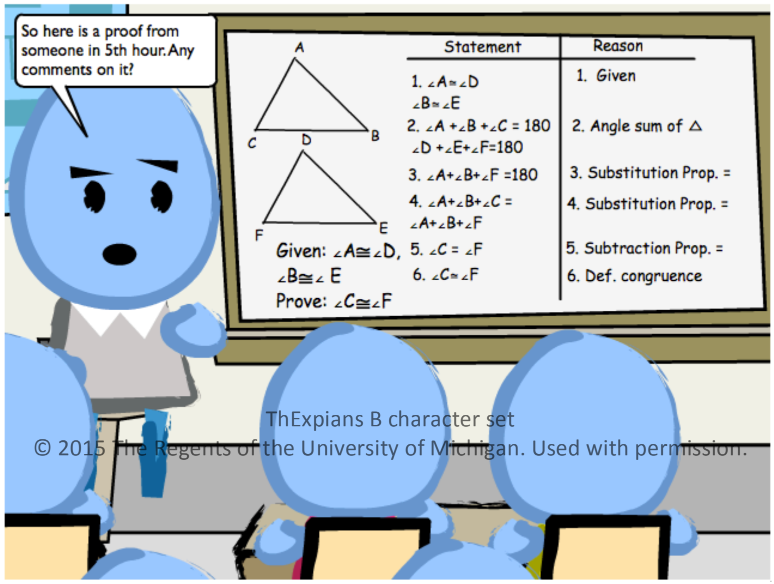

Figure 32: Problem used to target the *details* norm in storyboard 26002.

In storyboard 26002, the teacher begins by showing students a proof from "someone in 5th hour". The teacher's presentation of the proof and solicitation of comments from the class are meant to cue the proof checking ritual that was described in Chapter 3: students scrutinize the proof, then make comments—which the teacher mediates—about its various steps. A student says: "Step 2 says the angles sum to 180…but shouldn't it be the measures of the angles?" To which the teacher replies: "Oh sure, but we know what is meant. No need to dwell on little things like that. Other thoughts?" This storyboard thus depicts a breach of the norm because the teacher is prepared to cash a proof that does not explicitly maintain the angle/angle measure distinction. The control dual of this storyboard—26012—shows the same proof, with a student raising the same objection, only now, the teacher responds to the student comment by saying: "That's a great catch. We can't add or subtract angles, only their measures. So we actually need to fix some things." The teacher then proceeds to lead the class in rewriting the proof—specifically, the teacher adds a step that establishes the equal measures of the angles that are given to be congruent (definition of congruence), then rewrites the angle addition equation using $m\angle$[x] notation. The common ground across storyboards 26002 (breach) and 26012 (control) is a class discussion about the difference between the *substitution* and *subtraction* properties of equality, prompted by a student who confesses, “I always mix up substitution and subtraction.”

Review of the* less details *storyboards. Each of the *less details* storyboards depicts teacher’s actions that are *reasonable departures* from what usually occurs. In storyboard 26001, the congruence of the segments follows immediately from the definition of midpoint. That this implication is not explicitly stated reminds of the routine practice of leaving some of the details of any written record of a proof “for the reader.” In storyboard 26002, the notational error has no semantic bearing on the overall status of the argument; by forgiving this small detail, the teacher could be seen as conveying to the students that a mathematical argument is greater than the sum of its parts. Both storyboards thus show reasonable departures because being hyper-focused on minor details paints an overly ritualistic picture of what proof is actually like in mathematics (Harel & Sowder, 1998; Schoenfeld, 1988). The *less details* storyboards are complemented by the *more details* storyboards—storyboards that show teachers who request more details than what is normatively required when cashing a proof in an instructional exchange.

Developing storyboards that breach the norm by asking for more detail

The two storyboards that target the *more details* way of breaching the details norm were inspired by the fact that, while certain mathematical details seem to matter a great deal to whether a proof is accepted as valid currency in an instructional exchange, other details seem to matter less so. Chapter 3 describes an instance in episode CM-101802-5P-S4, where there is a discussion about the statements and warrants that are necessary to establish that two angles that form a linear pair are supplementary. In a different episode—MK-111102-3P-S3, also described in Chapter 3—the issue of the details that are necessary to warrant a claim of supplementarity from angles that are a linear pair is at stake in a similar way. The discussion about the appropriate level of unpacking of the claim that the measures of the angles that form a linear pair sum to 180 was a partial inspiration for storyboard 26003, the first storyboard that targets the *more details* norm (described below).

The second storyboard that targets the *more details* norm—26004—draws on the discussion about what can be assumed from a diagram that occurs in episode CM-101802-5P-S4. To recap the teacher states that proofs are always given with a diagram and that it is appropriate—perhaps even necessary—to make some assumptions about what is represented in the diagram in order to proceed with the proof. As described in Chapter 3, the lesson corpus shows instances of geometry teachers asking students to explicitly unpack some of the written statements in a proof into their more basic steps—the distinction between the supplements theorem and the definition of supplementary is an apt example. Both storyboards that target the *more details* norm probe whether participants would respond to episodes in which the teacher solicits this kind of unpacking for details that are normatively conveyed by the diagram.

Synopsis of storyboards 26003 and 26013. Figure 33 shows the problem that was depicted to target the *more details* norm in storyboard 26003.

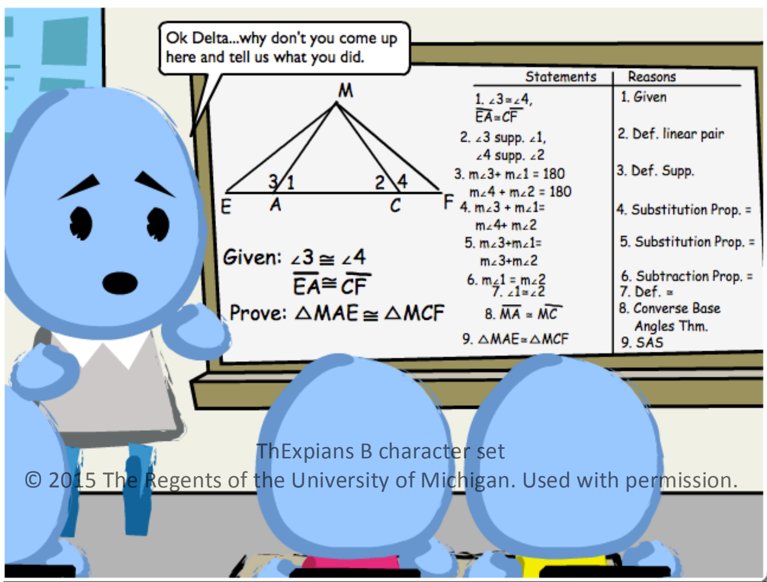

Figure 33: Problem used to target the *details* norm in storyboard 26003.

In storyboard 26003, the teacher calls a student, Delta, to the board to explain the proof. The student begins by saying that the first step was to write the givens then says that the next step was to show that angles 1, 3 and 2, 4 are supplementary, by the definition of linear pair. At this point in the storyboard, the teacher says to the student "Let's pause here for a moment." The teacher then goes into an explanation of why the student needs additional steps in the proof before saying that the angles are supplementary. In the breach version of the storyboard, the additional steps that the teacher insists on are a step that establishes the co-linearity of points E, A, C, and F, and a step that, as a result, establishes that the endpoints of rays AM and CM lie on line EF. Thus, the teacher insists that the students use the diagram to explicitly establish that angles 1,3 and 2,4 satisfy the definition of linear pair[38]. In the control version of this storyboard, the teacher still asks the student to provide an additional step—consistent with requests for clarification made by the teacher in episode CM-101802-5P-S4 and similar clarification made by a different teach in episode MK-111102-3P-S3—except now, all the teacher requires is for the

[38] A linear pair is a pair of angles formed when the endpoint of ray falls on a line (paraphrased from *Holt* Geometry, 2004, p. 28).

student to unpack step 2 into two different steps: one that identifies the angles as a linear pair (by definition of linear pair), then one that establishes them as supplementary, by the supplements theorem. The common ground across both storyboards is that, after explaining the additional steps that are needed in the proof, the teacher asks the students to add the steps to the proofs that are written on the board.

Synopsis of storyboards 26004 and 26014. Figure 34 shows the problem that was depicted to target the *more details* norm in storyboard 26004.

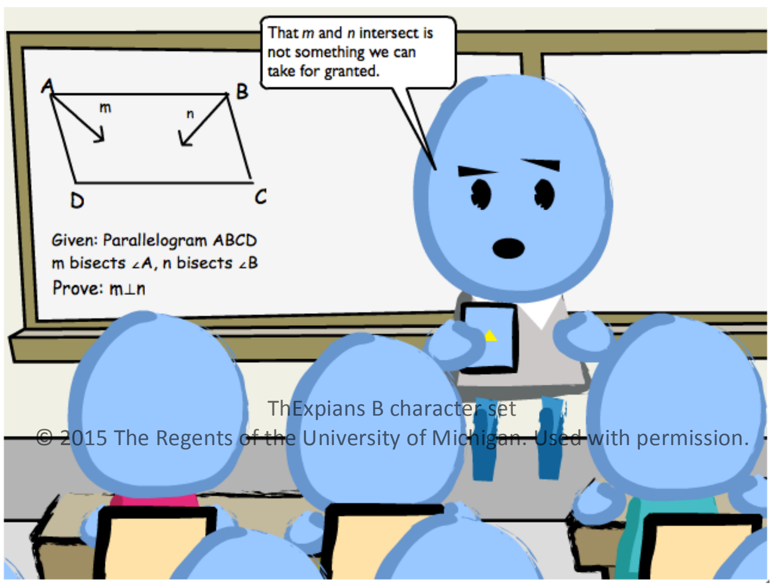

Figure 34: Problem used to target the *details* norm in storyboard 26004.

In storyboard 26004, the teacher is going over a proof problem from the book. The diagram shows a parallelogram with the angle bisectors of two consecutive angles. What is required is to show that the angle bisectors are perpendicular. In this storyboard, the teacher breaches the norm by problematizing the existence of a point of intersection for *m* and *n.* The teacher asks the students to make an argument that those two rays intersect before proceeding with the proof that they are perpendicular. This is a breach of the norm because the teacher is insisting on a detail—an argument that the rays will intersect at all, prior to establishing that they intersect at right angles—that is normatively conveyed tacitly by the diagram. In the control version of the storyboard, the teacher draws the diagram with the point of intersection marked (i.e., with a 'dot') and labelled with an *F*. Rather than problematize the existence of the point of intersection, the teacher asks students about a strategy for beginning the proof. Specifically, the teacher asks the students, "What do we do first" to which a student replies, "We write the givens" and the teacher follows up with, "Which is…?" A student then says: "We are given a parallelogram…and some congruent angles." The teacher cautions, "We need to be a little careful there. Besides a parallelogram, what are we actually given?"

The purpose of including this exchange in the control version of storyboard 26014 was to include an action that is consistent with the hypothesized norm for the kinds of details that are appropriate for a proof. In this exchange, the teacher could be seen as steering the student toward a more precise accounting of what is required for the proof, by insisting that conceptual entailments (e.g., an angle bisector implies congruent angles) are broken down into their more basic steps (e.g., a statement first that a ray is an angle bisector, followed by a statement of the congruence relations that follow from the definition of angle bisector). The common ground across each storyboard is a class discussion about the fact that the consecutive angles of a parallelogram are supplementary.

Review of the* More Details *storyboards. Each of the *more details* storyboards depicts a teacher enacting *reasonable departures* from what is hypothesized to normatively occur when checking proofs in geometry. In storyboard 26003, the teacher's actions could be seen as reasonable because asking students to provide more explicit statements and warrants is consistent with the kinds of teaching actions described in the *less details* section: It is routine for a geometry teacher to request students unpack proof statements into more basic parts. A similar argument warrants the reasonableness of the teacher's departure in 26004. The *less details* and *more details* storyboards probe teachers' recognition of the kind of work that is necessary for a proof to be cashed in the instructional exchange of doing proofs. The storyboards that target the *sequence norm* probe teachers' recognition of how semiotic resources are coordinated during the presentation of a proof.

Developing storyboards that breach the sequence norm

I developed six storyboards that target the *sequence norm*. These storyboards were inspired by classroom episodes (described in Chapter 3) that show students transcribing proofs from note sheets to classroom canvases. During the transcription, students generally work in silence, glancing from their note sheets to what they are writing on the board back to their note sheets every few seconds. A proof that is in the process of being transcribed is not required to cohere as a mathematical argument until the transcription is finished. This means, for instance, that a student transcriber might wait until the end of the transcription to draw or label a diagram that accompanies a proof—even though the steps in the proof referred explicitly to the diagram—or that a student might mark a diagram before writing the statements and reasons that warrant the markings.

A complete written record of a proof includes (1) a diagram that is (2) labelled and (3) marked, together with (4) a list of statements (including what is given and what is to be proved) and (5) reasons that warrant those statements. While there are finer-grained ways to describe the inscriptions that comprise a written record of a proof—one could, for instance, distinguish words from symbols in (4) or (5); or resolve (1) into strokes, dots, and other kinds of visual parts (Dimmel & Herbst, 2015)—these broad headings can be used to indicate the sequencing of semiotic resources when a proof is presented. For example, when a teacher presents a proof, it is routine for the teacher to begin by drawing and labelling the diagram that will accompany the proof; following this, the teacher will write the statements that are given and that which is to be proved, potentially marking the givens in the diagram once they have been written out. The teacher next might make a **T** chart on the board, to indicate where the *statements* (left side of the **T**) and *reasons* (right side of the **T**) will be written. The proof then proceeds in a coherent sequence, during which the teacher might describe an overall strategy for the proof before iterating a *statement-reason-marking* (where applicable, e.g., congruent segments could be marked once they have been proved, but a linear pair could not be[39]) protocol that fills in the **T** chart from left to right until the statement that is to be proved has been warranted by the previous steps. The teacher coordinates speaking with writing words and symbols and interacting with the diagram through gestures and markings. The range of semiotic resources is sequenced and coordinated by the teacher so that the in-progress proof is readable as it is being completed.

In contrast to the coherent sequencing of semiotic resources by teachers as they present proofs, when students transcribe proofs they could presumably reproduce (1)-(5) in any order. This hypothesis is based on student transcription practices that are recorded in the GRIP lesson corpus. The video episodes show students drawing and marking a diagram out of sequence; writing a proof out in its entirety before writing the statements that are given and that which is to be proved; and writing statements and reasons that refer to an unlabelled diagram. Even instances when student transcription work is not recorded, when students present a proof to the class, the presentation and narration of what they did always comes after the student has

[39] There are, however, resources (described in Chapter 3) through which the teacher could make the parts of a diagram more prominent, such as circling or tracing or gesturing at a specific part of the diagram to highlight the referent of a step in a poof.

produced a complete written record of the proof. That is, students do not seem to be expected to execute the kind of coordinated performance of a proof that is routine for teachers.

The availability of customizable graphic resources provides a means for designing storyboard probes—once again in breach and control pairs—that can gauge the extent to which teachers recognize the *sequence norm*. Stated as a hypothesis, the *sequence norm* holds that when transcribing a proof, students are not accountable to produce coherent sequences of inscriptions until the transcription is complete. Below, I provide summaries of the three sets of storyboards that target different aspects of the work of transcribing a proof. The first group of storyboards—27001 and 27002—show students that add markings to diagrams out of sequence. The second group of storyboards—27003 and 27004—show students that add labels to diagrams out of sequence. The third group of storyboards—27005 and 27006—show students that decouple statements from reasons when transcribing proofs. Following the summaries of the storyboards, I discuss how the teacher's actions could be seen as reasonable departures from the hypothesized norm.

Synopsis of storyboards 27001 and 27002. Each of these storyboards depicts a breach of the *sequence norm* by showing a teacher that takes issue with a proof in which the labels of a diagram are transcribed out of sequence. In storyboard 27001, the student transcribes all of the statements and reasons of the proof, but fails to label the diagram. This storyboard is inspired by episode MK-121702-3P-S3 from the GRIP lesson corpus, during which a student transcribes a proof, fails to label the diagram then returns to his seat. When it comes time for the student to explain what he did to the class, the teacher says to the student, "Can you put letters on yours?". That the student did not label the figure as he went yet still produced a set of statements and reasons that referred to the diagram is an indication that the student was transcribing, rather than performing, his proof at the board. The student was not deterred by the absence of diagrammatic referents for the statements and reasons he was writing on the board, because he already had established those referents on his paper, from which he was copying. That the teacher didn't require the student to undertake a more involved correction of the missing labels is an indication that the teacher does not expect students to perform proofs at the board in the same way that a teacher would be expected to perform a proof. In such a situation, the student is responsible for reproducing work that has already been done, rather than showing—via a performance of a proof—how such work is accomplished. Figure 35 shows a screenshot from storyboard 27001 that shows the student's finished transcription, accompanied by an unlabeled diagram.

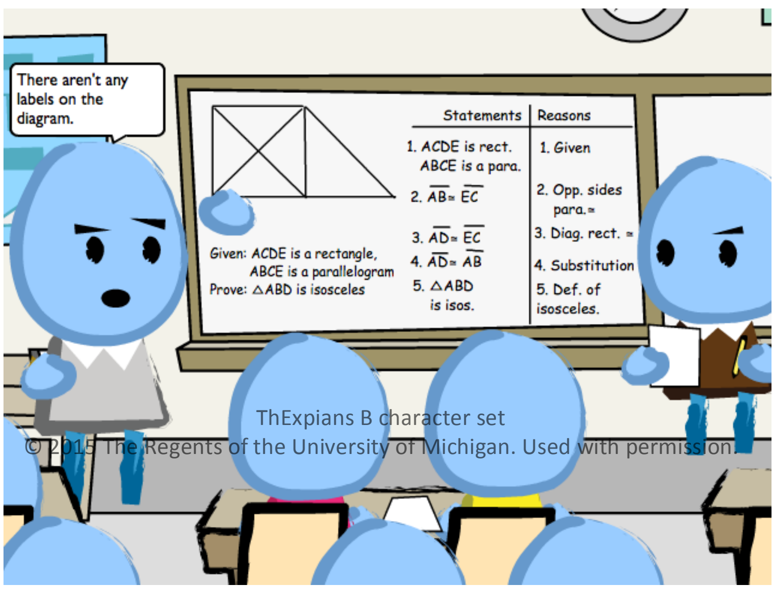

Figure 35: Problem used to target the *sequence* norm in storyboard 27001.

As opposed to simply allowing the student to label the diagram after the fact—and thereby accepting that the student has no responsibility for transcribing a proof in a coherent sequence—the teacher in storyboard 27001 breaches the norm by re-labelling the diagram: The teacher puts labels on the points that are different from those that are used by the student (Eta) in the written statements and reasons of the proof. The use of different labels for the diagrams requires the student to revise the proof. In the control dual of this storyboard—storyboard 27011—the teacher simply allows the student to apply the labels after-the-fact (as happens in episode MK-121702-3P-S3). The common ground across these storyboards is a student who shares a different method for completing the proof.

Storyboard 27002 also depicts a breach of the labels part of the sequence norm. In this storyboard, a student labels the points in the figure, but writes the proof using numeral labels for angles—e.g., angle 1, angle 3—before writing these numerals in the diagram. As the student—Beta—completes the transcription, a different student—Mu—asks the teacher "which angles are 2 and 4…and 1 and 3". The teacher breaches the *sequence norm* in storyboard 27002 by saying, "That's a good question" then telling Beta that, "We can't follow what you are doing because you haven't labeled the angles in the diagram." The teacher asks Beta to "erase those two steps, label the angles in the diagram, and then try again." Figure 36 shows Beta's work in the proof before the breach of the norm occurs.

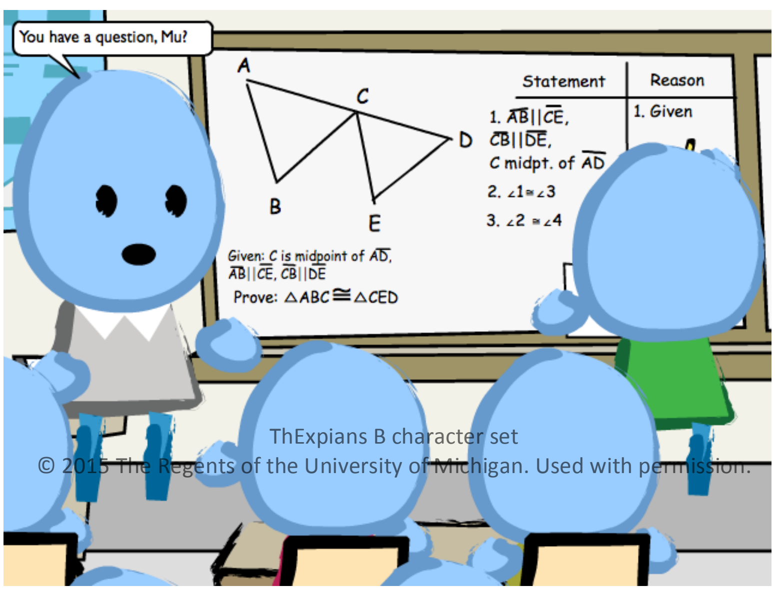

Figure 36: Problem used to target the *sequence* norm in storyboard 27002

The teacher in the control dual of this storyboard—storyboard 27012—complies with the sequence norm by telling Mu to wait until Beta is finished before asking questions. Thus, in the control, the teacher effectively affirms that the statements and reasons of the proof are not required to cohere until the written record of the proof is completed. The common ground across both storyboards is a discussion between two students about the nature of proof and whether two proofs that use different evidence could both be correct. Specifically, one of the students argues that a proof with more steps has more evidence and is therefore better than a shorter proof.

Synopsis of storyboards 27003 and 27004. Each of these storyboards depicts a breach of the *sequence norm* by showing a teacher that takes issue with a proof in which the markings of a diagram are transcribed out of sequence. Narratively, these storyboard pairs are comparable to storyboards 27001 and 27002 (and looking ahead: 27005 and 27006). In storyboard 27003, a student is in the process of transcribing a proof and marks all of the congruence relations in the diagram before establishing those relationships in the written statements and reasons of the proof. Two other students in the class say to the teacher, “I don’t know see how Gamma got angle 1 and angle 2 congruent”, and “Or angle 3 and angle 4…isn’t that the prove?”. In the breach version of the storyboard—similar to what happens in storyboard 27002—the teacher takes up the student comments then asks Gamma to erase the markings in the diagram and add them back in only after those relationships had been established in the proof. Figure 37 shows a screenshot of the proof used to represent the breach.

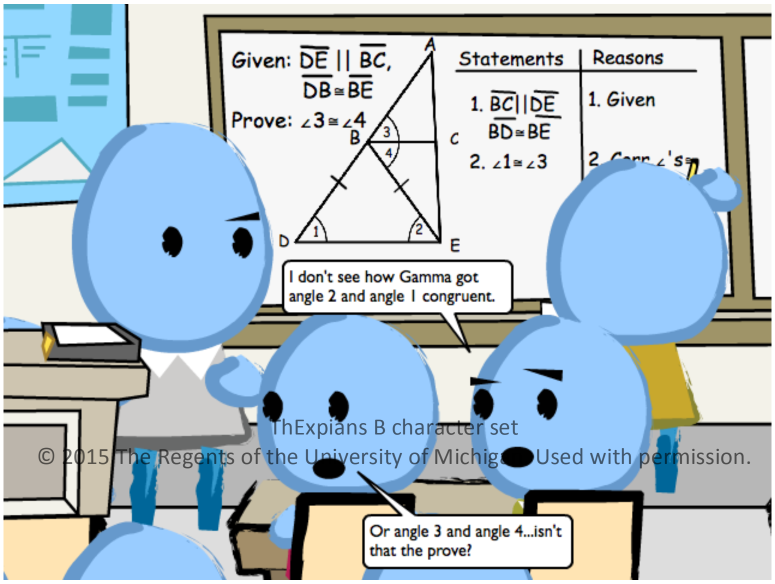

Figure 37: Problem used to target the *sequence* norm in storyboard 27003.

In the control dual of storyboard 27003, the teacher tells the students that ask about the markings, “Gamma is not finished yet. You will be able to ask questions when Gamma is finished.” This response conforms to the hypothesized norm that a student transcription of a proof need not cohere as an argument until the transcription is completed. The *common ground* across both storyboards is a student question about how to identify the appropriate transversals when making arguments about corresponding and alternate interior angles.

In storyboard 27004, a student completes the transcription of a proof without marking the diagram then sits down, similar to storyboard 27001, in which the student sits down without labelling the diagram. Rather than allow the student to come to the board and apply the labels all at once after the fact, the teacher breaches the *sequence norm* by insisting that the student redo the proof, this time marking the properties that hold in the diagram as they are shown to be true. Figure 39 shows a screenshot of the proof used for this storyboard.

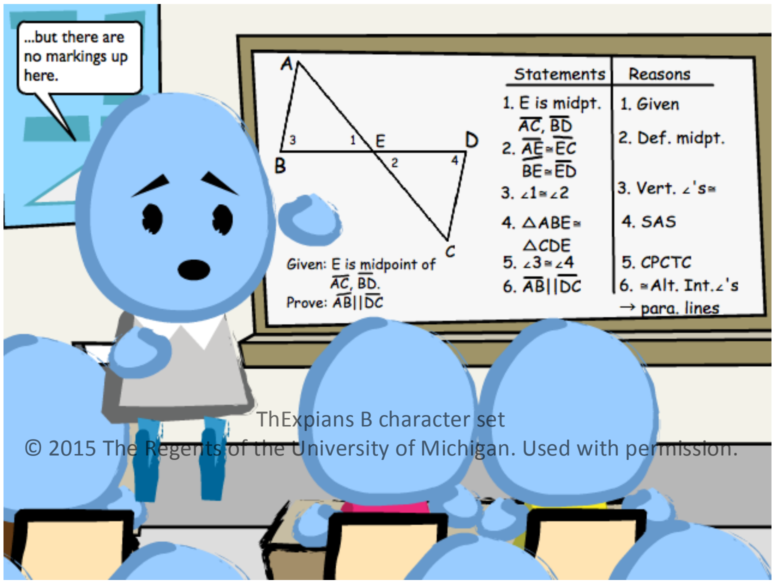

Figure 38: Problem used to target the *sequence* norm in storyboard 27004.

In the control dual of the storyboard, the teacher permits the student to apply the labels to the diagram without redoing the steps in the proof. This action conforms to the hypothesized norm that the sequence in which the different parts of a proof are transcribed has no bearing on the validity of the proof. What matters is that the final written record of the proof is complete—not the order in which its parts were completed. The *common ground* across both storyboards is the teacher's observation that the student has left the markings off the diagram.

Synopsis of storyboards 27005 and 27006. Each of these storyboards depicts a breach of the *sequence norm* by showing a teacher that takes issue with a proof in which the reasons of a proof are transcribed out of sequence. In storyboard 27005, a student completes the transcription of the proof, first writing all of the statements then writing all of the reasons. In the process of doing so, the student omits a reason for one of the statements in the proof. The teacher notices this omission. To breach the norm, the teacher insists that the student redo the proof from the point where the reason was omitted. Figure 39 shows a screenshot of the proof used in storyboard 27005.

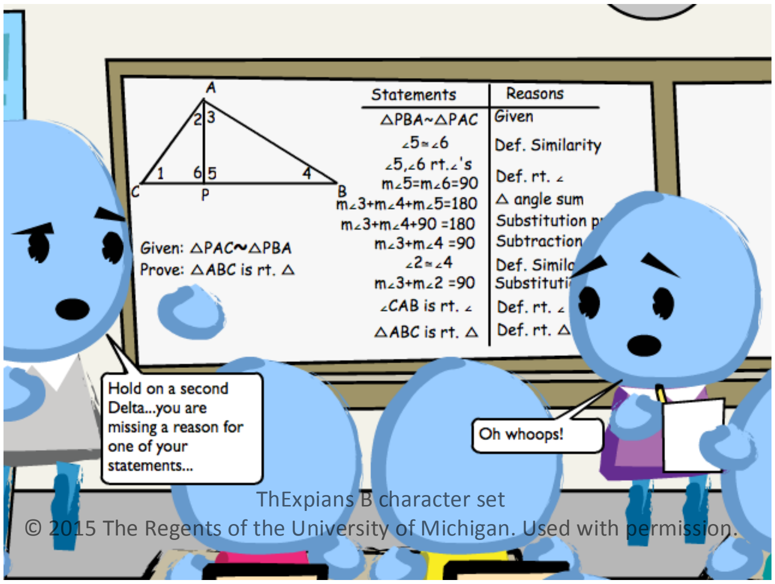

Figure 39: Problem used to target the *sequence* norm in storyboard 27005.

In both the breach and control versions of the storyboard, the student that omits the reason in the proof is able to identify what the missing reason is. In the control version of the storyboard, the teacher permits the student to insert the missing reason into the proof without altering any of the rest of it. This action conforms to the hypothesized norm that a proof need not cohere until its transcription is completed. The *common ground* across both storyboards is a student who volunteers a different strategy for doing the proof.

In storyboard 27006, a student transcribes a proof by first transcribing all of the statements of the proof then copying the reasons. As the student is completing the transcription, a different student in the class asks: "How did Omicron get that AD and AB are congruent?" That is, the student is asking for the reason that justifies a step that Omicron has already written. The teacher breaches the *sequence norm* by asking the student to erase the unjustified steps and redo the proof, writing reasons for each statement in sequence. Figure 40 shows a screenshot of the proof that was used in this storyboard.

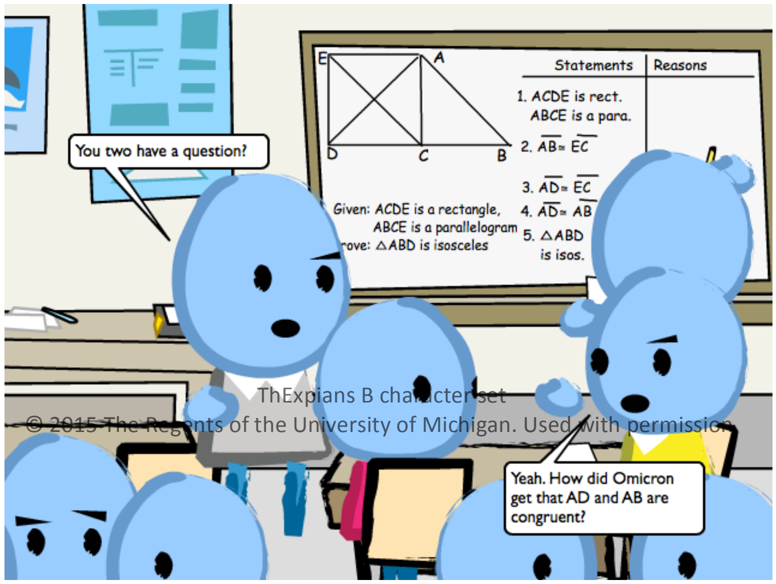

Figure 40: Problem used to target the *sequence* norm in storyboard 27006

In the control dual of storyboard 27006—similar to the control dual of storyboard 27003 and 27002—the teacher says to the student that asks the question: "Omicron is not finished yet. You will have a chance to ask question when she is done writing hers on the board." This action complies with the hypothesized norm that the student transcribing work is allowed to do so in any manner that makes that transcription convenient. The work is not to be checked or cashed until the transcription is completed. The *common ground* across both storyboards is a conversation between two students about the difference between the definition of an isosceles triangle and the converse of the base angles theorem.

Review of the sequence norm storyboards. The storyboards that breach the *sequence norm* depict teachers that interfere with the routine ways that proofs are transcribed in geometry classrooms. In each of these storyboards, the teacher breaches the *sequence norm* by attempting to make the proof the student is transcribing accountable as a coherent mathematical argument before the transcription is completed. These deviations on the part of the teacher from what is expected to be the normative ways that proofs are transcribed in geometry classrooms could be defended on the grounds that, by insisting that student arguments cohere as they are being produced, the teacher is creating opportunities for students to engage in the mathematical practices associated with justifying and explaining one's work. After all, when a teacher presents proof, there is certainly an expectation that the argument will be developed in a sequence that is coherent as it is being completed—a teacher would not be able to transcribe an argument in any convenient order then discuss it only after the transcription was completed. This is to say that the teachers in these storyboards are not acting punitively toward their students. Their requests that the students not merely transcribe but redevelop the proof in a coherent sequence are mathematically defensible and hence constitute *reasonable departures.*

I have described the 10 storyboard probes that were designed to target different aspects of the *details* and *sequence* semiotic norms. The key feature of the storyboard probes I created for this study is the matching of breach storyboards to control storyboards, so that the two different versions of the storyboard are minimally different from each other. The possibility to breach/control storyboard pairs was an affordance of the customizable graphic language that was used to create the storyboards. The final section of this chapter describes how these storyboards were used in a multimedia survey instrument that was designed to optimize comparisons of responses across and within different experimental conditions.

Instrumentation: creating a multimedia questionnaire

The 10 pairs of storyboards described above were used as probes in a multimedia survey instrument. The 20 storyboards—10 breach storyboards and their 10 control duals—were used to create 5 experimental groups. Each group contained 2 breach storyboards two storyboards that represented different ways of breaching the same norm and two storyboards that breached no norms (their storylines matched those of different breach storyboards). Figure 41 shows the assignment of storyboards to experimental groups.

Details norm		Sequence Norm		
Group 1	Group 2	Group 3	Group 4	Group 5
26001	26011			
26002	26012			
	26003	26013		
	26004	26014		
		27001	27011	
		27002	27012	
			27003	27013
			27004	27014
27015				27005
27016				27006

Figure 41: Assignment of storyboards to experimental groups.

The 5 experimental groups are the columns of Figure 41. The rows of Figure 41 show the pairing of breach storyboards with their control duals. For example, the first column of Figure 41 displays the 4 storyboards that were used in the first experimental group. Storyboards 26001 and 26002 depict breaches of the *less details* norm, while storyboards 27015 and 27016 are the control duals of storyboards that depict breaches of the reasons aspect of the *sequence norm*. The first row of Figure 41 shows the connection between storyboard 26001 and its control dual—assigned to a different experimental group—storyboard 26011. Thus, each participant in the study viewed and answered questions about two breach storyboards and two control storyboards. This design enabled two categories of planned comparisons that are discussed in Chapter 5: Inter-group comparisons across experimental conditions (i.e., breach vs. control) and intra-group comparisons within experimental groups.

The structure of the instrument was the same for all storyboards across all groups: participants were shown a classroom storyboard then asked a series of open- and closed-ended follow up questions. The open-ended questions included a general open-response prompt—What did you see happening in this scenario?—and "please explain your rating" follow up prompts to each of the closed-ended rating questions. The phrasing of the general open-response prompt and also the open-response follow-up to the closed-ended rating questions was based on previous instruments that used multimedia surveys to conduct virtual breaching experiments (Herbst, Aaron, Dimmel, & Erickson, 2013).

Participants were asked three closed-ended rating questions and two closed-ended multiple-choice questions. The rating questions used a 6-point Likert-like response format to rate appropriateness that ranged from "very inappropriate" (1) to "very appropriate" (6). For all three rating questions, participants were asked to rate different aspects of the teacher's work in each storyboard they viewed. The multiple-choice questions were designed to provide grounds on which teachers might justify or criticize the way that the teacher reviewed the proof in each storyboard. Participants were asked to complete the following sentences: "To justify the way the teacher reviewed the work on the board, I would say the teacher…" and "To criticize the way the teacher reviewed the work on the board, I would say the teacher…" Each of these questions was followed by the same five choices that offered different grounds on which the teacher's actions could be justified or criticized. The justification/criticism question and choices were based on questions used in previous virtual breaching experiments (Herbst, Aaron, Dimmel, & Erickson, 2013). The five available choices included four choices that were linked to different stakeholders to which mathematics teachers are obligated in their work as professionals (Herbst & Chazan, 2012) and one choice that was an escape clause, i.e., "I would not justify/criticize what the teacher did". Comparisons across and within conditions of participant responses to these questions are reported in Chapter 5.

In addition to the general rating question, each participant was asked two rating questions that targeted specific moments in each storyboard. These questions were designed to capitalize on the minimal differences between each breaching storyboard and its control dual. One of the targeted rating questions zoomed into the moment in the storyboard where the teacher either: (1) breaches or (2) complies with the target semiotic norm. This “zooming in” was achieved by showing participants a segment of the storyboard. Such segments were 3 to 5 frames long, matched across breach/control pairs (so, a 3 frame clip in a breaching storyboard was matched with a 3 slide clip of its corresponding control storyboard).

The purpose of this targeted rating question was to focus participants’ ratings on the part of the storyboard where the teacher complies with or departs from the norm. These targeted rating questions came after the general rating question described above. In addition to a targeted rating question that focused on the point in the storyboard where the teacher complies with or departs form a target norm, participants were asked an additional targeted rating question. This additional question focused on the *common ground* across each breach/control storyboard pair. For this other targeted rating question—what I call the *targeted distracter* rating question—participants in the breach/control conditions were shown identical sets of 3-5 images in which the teacher in the storyboard completes an instructional action unrelated to the target norm. It was possible to identify such segments because each set of breach/control storyboards were identical except during those parts of the storyboard that represent the breach of (or compliance with) the target norm. The purpose of including the two types of targeted rating questions was to enable comparisons across the breach and control conditions. These comparisons are reported in Chapter 5.

One final note on the design of the targeted rating and targeted distracter questions. The order in which participants would answer either of the targeted rating questions was varied across the storyboards. For 5 breach/control pairs, participants answered the targeted breach/control rating question before responding to the targeted distracter question. For the remaining 5 storyboards the opposite was true. The order of the questions was matched across breach/control pairs, so that, for instance, if the targeted breach rating questions appeared first in storyboard 26001, the targeted control rating question appeared first in storyboard 26011 (its control dual). A coin flip was used to determine which of the 5 storyboard pairs would display the targeted breach/control rating first. The purpose of varying the order to these questions was to control for any effect that might come from which of the questions was asked first.

Each of the 5 groups described in Figure 41 had 6 variations, giving a total of 30 experimental conditions. These were variations of the order in which the participants viewed and answered questions about the 4 storyboards in each condition. There are 24 possible combinations in which the storyboards in a condition could appear. These 24 possibilities were reduced to 6, on the grounds that the differences between which of the breach or control storyboards appeared were marginal when compared to whether a participant viewed two breach and two control storyboards, or a breach storyboard then a control storyboard, etc. Figure 42 shows the 6 orderings that were used across each of the 5 groups of storyboards to create 30 experimental conditions.

Order 1	**Order 2**	**Order 3**	**Order 4**	**Order 5**	**Order 6**
Breach	Breach	Breach	Control	Control	Control
Breach	Control	Control	Breach	Breach	Control
Control	Breach	Control	Control	Breach	Breach
Control	Control	Breach	Breach	Control	Breach

Figure 42: The 6 orderings of Breach (B) and control (C) storyboards that were used for each of the 5 experiment groups

Participants were randomly assigned to conditions in two ways. First, a participant was randomly assigned to one of the five groups of storyboards. The random assignment was conducted so that the numbers of participants assigned to each storyboard group was balanced. Once participants were assigned to a group, they were then randomly assigned to one of the 6 orderings of that group reported in . Once again, the assignment of participants to conditions within groups was conducted so as to balance the number of participants assigned to each condition. Chapter 5 describes the piloting of the instrument with a sample of 73 secondary teachers.

Review of the Virtual Breaching Experiment with Control Instrument

The method for gathering data explained above is a *virtual breaching experiment with control.* The key innovations of this method are the use of paired, minimally different breach/control storyboards that are viewed by participants in different experimental conditions and the use of targeted rating questions that zoom in on specific teaching actions in each storyboard. The goal of using minimally different storyboards and different types of targeted rating questions is to warrant comparisons of responses within and across experimental conditions. Five experimental groups derived from the 10 breach/control storyboard pairs (20 storyboards total). Each condition included 2 storyboards that breach a norm and 2 storyboards that do not breach a norm, for a total of 4 storyboards per condition. The groups were defined to optimize within group and across condition comparisons of participant responses to open- and closed-ended questions. Chapter 5 reports on the analysis of responses to closed-ended questions, while Chapter 6 reports on the analysis of responses to open-ended questions.

CHAPTER 5:

DO TEACHERS RECOGNIZE BREACHES AS APPROPRIATE? AN ANALYSIS OF CLOSED-ENDED RESPONSES.

I report here the analysis of closed-response data collected with the instrument described in Chapter 4. There were two types of closed-response data gathered for this study. The first type of data is participants' responses to the three closed-ended rating questions: one general rating question and two targeted rating questions, each of which prompts a participant to rate the appropriateness of the teacher's work a scenario. A second type of data is participants' responses to the two multiple-choice, sentence completion questions: (1) "To justify the way the teacher reviewed the work on the board, I would say the teacher…" and (2) "To criticize the way the teacher reviewed the work on the board, I would say the teacher…". The rating questions were analyzed within and compared across experimental groups. These comparisons were conducted using independent (across) and paired (within) samples *t*-tests. The responses to the multiple-choice questions were analyzed using binomial tests.

The chapter begins by providing summary information about the 73 teachers in the sample. Following this overview of the sample, I state the hypotheses about the differences in mean score on the various rating questions and report the results of the statistics that were used to test these hypotheses. The final part of the chapter reports the analysis of the responses to the multiple-choice questions. The coding and analysis of the open-response data is reported in Chapter 6.

Data collection for the virtual breaching experiment

Data was gathered from 73 secondary teachers over the course of 4 data collection periods during the 2013-2014 academic year. Participants were recruited from teachers living in South East Michigan via email during spring and summer of 2013[40]. Participants completed the virtual breaching experiment with control instrument as part of a suite of instruments that were deployed online using the Lesson*Sketch* (www.lessonsketch.org) environment. The additional instruments that participants completed during the data collection are not part of this study. Participants were also administered a background questionnaire that gathered personal (e.g., gender, years of teaching experience) and demographic information on all of the participants in the sample.

The 73 secondary teachers that participated in the pilot studies were current teachers at high schools in South East Michigan. 55 of the 73 teachers had taught high school geometry for at least 1 year. I refer to this subset of the sample as *geometry teachers.* Of these teachers that had taught geometry, 39 had more than 3 years of experience as a high school geometry teacher. The mean number of years of experience teaching geometry for the geometry teachers in the sample was 6.85 years of experience, with a standard deviation of 6.016 and a mode of 4. The maximum number of years of experience teaching geometry was 31 years.

Participant management protocol

During three of the data gathering sessions—one in October (n =17), one in November (n = 29), and one in January (n =11)—participants completed the instruments using University of Michigan provided[41] laptop computers during a daylong, in-person session for which I was the

[40]Participant recruitment was managed by the GRIP research group with the support of NSF grant DRL- 0918425 to principal investigator Patricio Herbst. Neither the instrument used nor this analysis necessarily represent the views of the Foundation.
[41] These computers were loaned to the GRIP research project by the School of Social Work.

in-room facilitator[42]. The final data collection was conducted remotely, with participants (n = 16) using their own computers to complete the instruments online during a 10-day window in March[43]. During the online data collection window, I played the role of facilitator and participant manager. Participants who completed the instrument online were already familiar with the Lesson*Sketch* environment, having completed an in-person pilot of different instruments in August 2013. Participants were paid a $100 honorarium for the in-person sessions and a $50 honorarium for the shorter, online session. Participants were paid for time they dedicated to completing all of the instruments that were used during the in-person (6 hours) and online (3 hours) data gathering collections sessions.

My purpose as facilitator was to explain to participants what they would be doing over the course of the day and to be available in the event that participants encountered technical difficulties logging into the system for completing the instruments. I also served as a de facto proctor to enforce a quiet environment. Discouraging talking during the data collection periods was important to help ensure that participants were providing individual responses, as opposed to discussing the various questions with their colleagues. Given the matched breach/control design of the experiment, maintaining a quiet working room also helped to ensure that participants in different conditions were not aware of alternate versions of the storyboards they were viewing.

Beyond providing an implicit presence to establish a quiet workspace for participants, as facilitator I also explicitly stated—during my welcoming remarks at the beginning of each in-person session—that participants were to complete the various experiences they were assigned

[42] I acknowledge the support of Pat Herbst and colleagues at the GRIP research lab—including Nicolas Boileau, Amanda Milewski, Inah Ko, Vu Minh Chieu, Kristi Hanby, and Ander Erickson—who worked in other capacities to facilitate these data collection periods.
[43] Data gathered during the online pilot was supported by a Rackham research grant. I acknowledge assistance from Kristi Hanby in recruiting, managing, and compensating participants during this part of the study.

individually. I also explicitly stated that if participants had questions they should ask me, rather than another participant. I explained that I would not be able to answer questions about the substance or interpretation of any of the instruments, but that I would help with technical issues people encountered. Following this participant-interaction protocol was a means of ensuring the integrity of the data gathered during the data collection. There were no participant management issues during the in-person or online data collection periods.

Assignment of participants to experimental conditions

The five experimental groups described in Chapter 4 and summarized in Table 1 (below) generated 30 online experiences: 6 different orderings of the breach/control storyboards for each experimental group. Because I did not know ahead of time which of the participants that had agreed to attend the in-person sessions would, in fact, show up, participants were randomly assigned to a storyboard ordering within a group based on the order in which they checked in to an in-person session. Mimicking this procedure for the online pilot, participants were assigned to groups based on the order in which they responded via email to the online solicitation.

Table 1: Summary of experiment groups, with sample sizes and storyboards

n=16 Group 1	*n*=13 Group 2	*n*=15 Group 3	*n*=14 Group 4	*n*=15 Group 5
Less Details B1	More Details B1	Labels B1	Markings B1	Reasons B1
Less Details B2	More Details B2	Labels B2	Markings B2	Reasons B2
Reasons C1	Less Details C1	More Details C1	Labels C1	Markings C1
Reasons C2	Less Details C2	More Details C2	Labels C2	Markings C2

Note: Bx and Cx refer to the breach/control storyboards in a condition; the table shows which storyboards were in each group. Each experiment group had 6 different orderings of the breach and control storyboards within the group (described below).

The assignment of attendees to experiences was completed using a 2-part randomization procedure. I indexed the 30 experiences with a two-digit number, the first of which recorded the experiment group (1-5) and the second of which recorded its arrangement of storyboards (1-6). For example, experience 13 was the third arrangement (Breach, Control, Control, Breach) of storyboards that were assigned to group 1. I then generated 6 random orderings[44] of the five experiment groups. This yielded a list with 30 total entries—in blocks of {1, 2, 3, 4, 5}—such that the first block of 5 was the first random ordering of experiment groups, the second block of 5 was the second random ordering of groups, etc. There were thus 6 placeholders—one in each of the 5 blocks—for each experiment group throughout the list. This was the first part of the procedure.

For the second part of the procedure, I generated 5 random orderings of the 6 within-group storyboard arrangements. The first ordering was for the arrangements of storyboards within group 1, the second ordering was for the arrangement of storyboards within group 2, etc. I then used these 6 random orderings to fill in the placeholders in the master list. For example, if {6, 2, 3, 4, 1, 5} were the random order generated for storyboard arrangements in group 1, then this would mean that the placeholders in the master list for group 1 were replaced with 16, 12, 13, 14, 11, and 15—i.e., the first participant assigned to group 1 would be assigned to arrangement 6 of its component storyboards, the second participant would be assigned to arrangement 2, etc. This method of assigning participants to experiences helped to ensure both balanced numbers of participants assigned to groups and, within those groups, balanced assignment to storyboard arrangements.

Even given these steps, it was not possible to achieve perfect balance across the conditions and arrangements, on account of factors that were beyond my control. For example, during the January data collection period, 2 participants that were assigned experiences did not complete them. For the online data collection sessions, there were 7 participants that were

[44] I used the random sequence generator at www.random.org.

assigned experiences that did not complete them. Participant data from incomplete experiences was not part of the sample.

The next section reports the analysis of participant responses to the rating questions. The analysis of participants' open responses is reported in Chapter 6. This section begins with a descriptive overview of the responses to each of the three rating questions participants answered following each storyboard. Descriptive statistics are reported for each experiment group, stratified by rating-question type. Following the reporting of descriptive statistics, I define the measures that were derived form the rating response data, and use these measure to state the hypotheses that the virtual breaching experiment with control was designed to test. The final part of the next section reports the results of the significance testing.

Analysis of the responses to the rating questions

Descriptive overview of responses to the rating questions

The purpose of including closed-response rating questions in the instrument was to gauge the extent to which participants reacted to storyboards in which the teacher was shown to breach a hypothesized semiotic norm. The closed response-rating questions were designed to probe for reactions at two levels. The general rating question—How appropriate was the teacher's review of the proof?—was designed to gauge a participant's overall reaction to the storyboard. This was the first rating question that participants answered. From the responses to this first rating question, I defined an *episode appropriateness* (EA) measure from the means of participants' responses to parallel items (breach with breach, control with control) within each experiment group.

The targeted rating questions were designed to gauge a participant's reaction to the breach of the norm in a more focused way than the general rating question. One segment of the storyboard that participants rated for the targeted rating question was the segment of the storyboard during which the teacher breaches (treatment) or does not breach (control) the target norm. Responses to the *segment of interest appropriateness* ($S_{(I)}A$) rating questions were used to define a segment of interest appropriateness measure from the means of participant responses to parallel items (breach with breach, control with control) within each experiment group. Participants also rated the segment of the storyboard during which the teacher completes some other action that occurs in each treatment/control episode pair. Responses to the *distracter segment appropriateness* ($S_{(D)}A$) rating questions were used to define a distracter segment appropriateness measure from the means of participant responses to parallel items (breach with breach, control with control) within each experiment group.

The EA, $S_{(I)}A$, and $S_{(D)}A$ measures were derived from the means of ratings on corresponding storyboards within each experiment group. Within each experiment group, control storyboards corresponded with control storyboards and breach storyboards corresponded with breach storyboards. For example, EA(Group 1)$_{treatment}$ was the mean of the EA ratings on the 2 treatment storyboards in experiment group 1 Likewise, EA(Group 1)$_{control}$ was the mean of the EA ratings on the 2 control storyboards in experiment group 1. In each experiment group, the treatment storyboards were the storyboards that show a teacher breaching a hypothesized semiotic norm. The control storyboards in each experiment group were the storyboards that show a teacher not breaching any hypothesized norms. The EA, $S_{(I)}A$, and $S_{(D)}A$ measures were used for the hypothesis testing that is reported below.

Participants answered the EA, $S_{(I)}A$, and $S_{(D)}A$ questions after each storyboard they were shown. As described above, within each experiment group, participants viewed two breach storyboards and two control storyboards. Below, I use box-and-whisker plots to represent participant responses to each type of rating question within each experiment group. In addition to the box-and-whisker plots, I report means and standard deviations by question type within each experiment group. The reports of the descriptive statistics in the next section are just advances on the data. I hold discussion of the data and drawing conclusions from the data until I report the statistics I used to test hypotheses about participants reactions to the different storyboard conditions.

Descriptive statistics for experiment group 1

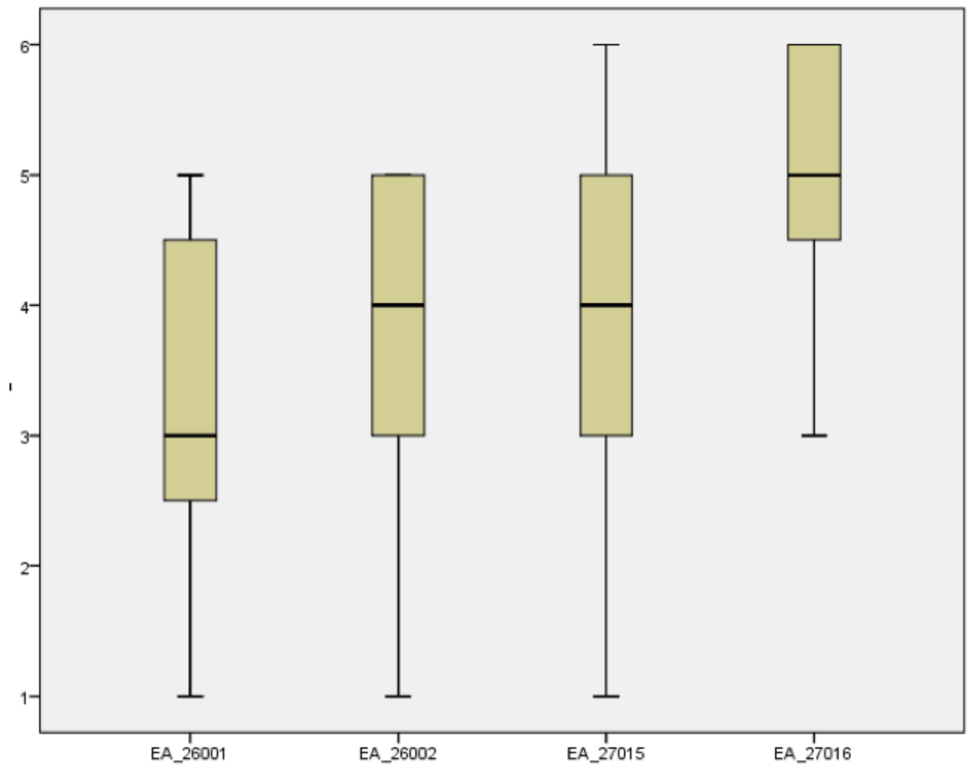

Figure 43: Box-and-whiskers plots for responses to the *episode appropriateness* rating questions for experiment Group 1 (*n* =16).

Figure 43 shows box-and-whiskers plots for responses to the *episode appropriateness* rating questions to the four storyboards in Group 1. Sixteen participants viewed and answered questions about each of these four storyboards. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the response to the first rating question indicate that the responses to the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for the two control storyboards. The means and standard deviations for the *episode appropriateness* ratings are reported in Table 2.

Table 2: Mean and standard deviation for the 4 *episode appropriateness* questions answered by Group 1 participants.

EA_(Storybaord) (*n*=16)	**Mean**	**Standard Deviation**
EA_26001	3.31	1.25
EA_26002	3.63	1.26
EA_27015	4.13	1.45
EA_27016	4.94	1.12

Table 2 reports the means and standard deviations of responses to the *episode appropriateness rating* questions for Group 1. The EA(Group_1)$_{treatment}$ measure was derived from the means of EA_26001 and EA_26002. The EA(Group 1)$_{control}$ measure was derived from the means of EA_27015 and EA_27016. Results of significance testing for across-condition and within-group differences in means are reported below.

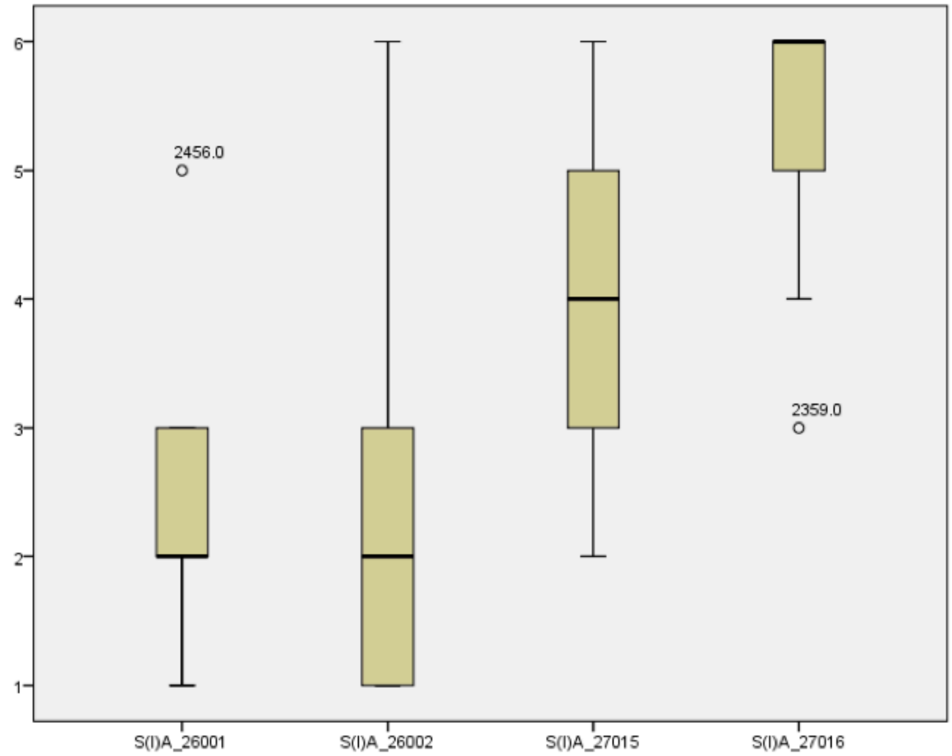

Figure 44: Box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions for experiment Group 1 (*n* =16).

Figure 44shows box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions to the four storyboards in Group 1. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The two points marked with open circles and labeled with 4-digit numbers are mild outliers[45]; the 4-digit numbers are participant IDs. The plots of the responses to the *segment of interest* rating question indicate that the responses to each of the breach storyboards had median, first-quartile, and third-quartile scores that were less than the median, first-quartile, and

[45] A *mild outlier* is a data point that is between the *inner* and *outer* fences of the plot. The *inner fence* has a lower bound of (Q3-(1.5*IQR)) and an upper bound of (Q1+(1.5*IQR)). The *outer fence* has a lower bound of (Q3-(3*IQR)) and an upper bound of (Q1+(3*IQR)). In these definitions, IQR is the inter-quartile range.

third-quartile scores for each of the control storyboards. The means and standard deviations for the *segment of interest appropriateness* ratings are reported in Table 3

Table 3: Mean and standard deviation for the 4 *segment of interest appropriateness* questions answered by Group 1 participants.

$S_{(I)}$A_**(Storybaord)** **(*n*=16)**	**Mean**	**Standard Deviation**
$S_{(I)}$A _26001	2.31	0.95
$S_{(I)}$A _26002	2.31	1.45
$S_{(I)}$A _27015	4.13	1.36
$S_{(I)}$A _27016	5.44	0.89

Table 3 reports the means and standard deviations for the *segment of interest appropriateness* rating questions for Group 1. The *segment of interest appropriateness* measures for Group 1 were derived from the means of the corresponding storyboards, consistent with the procedure described above. The $S_{(I)}$A(Group 1)$_{treatment}$ measure was derived from the means of $S_{(I)}$A_26001 and $S_{(I)}$A_26002. The $S_{(I)}$A (Group 1)$_{control}$ measure was derived from the means of $S_{(I)}$A_27015 and $S_{(I)}$A_27016. Results of significance testing for across-condition and within-group differences in means are reported below.

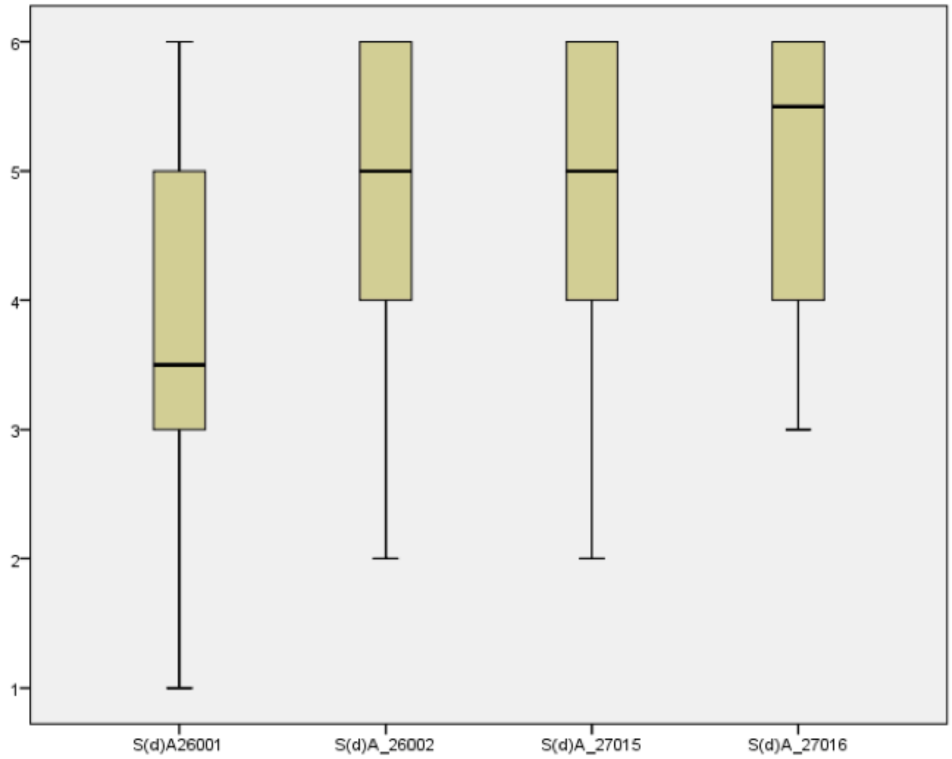

Figure 45: Box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions for experiment Group 1 (n =16).

Figure 45 shows box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions to the four storyboards in Group 1. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. By contrast with the plots of the data for the *segment of interest appropriateness* rating question, the plots of the responses to the *distracter segment* rating question indicate more balanced reactions to the segment of the storyboard in which the teacher does an action that is unrelated to the norm. The means and standard deviations for the *distracter segment appropriateness* ratings are reported in Table 4.

Table 4: Mean and standard deviation for the 4 *distracter segment appropriateness* questions answered by Group 1 participants.

$S_{(D)}$A_**(Storybaord) (*n*=16)**	**Mean**	**Standard Deviation**
$S_{(D)}$A _26001	3.56	1.46
$S_{(D)}$A _26002	4.88	1.09
$S_{(D)}$A _27015	4.75	1.29
$S_{(D)}$A _27016	5.00	1.15

Table 4 shows the means and standard deviations of participants' responses to the *distracter segment* appropriateness rating questions in Group 1. The *distracter segment appropriateness* measures for Group 1 were derived from the means of the corresponding storyboards. The $S_{(D)}$A(Group 1)$_{treatment}$ measure was derived from the means of $S_{(D)}$A_26001 and $S_{(D)}$A_26002. The $S_{(D)}$A(Group_1)$_{control}$ measure was derived from the means of $S_{(D)}$A_27015 and $S_{(D)}$A_27016.

Descriptive statistics for experiment group 2.

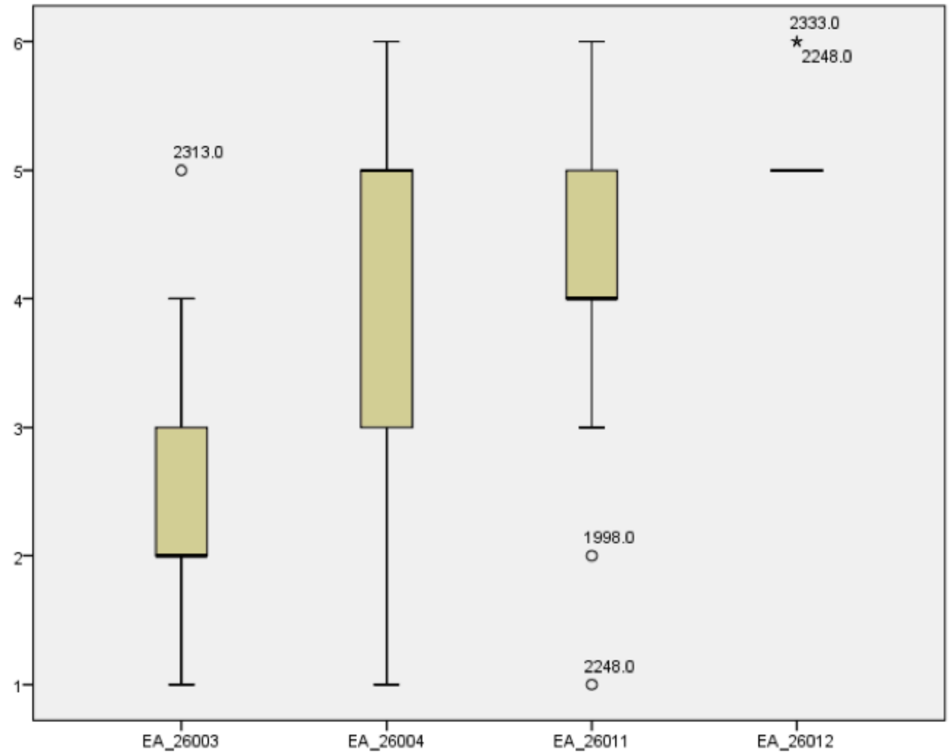

Figure 46: Box-and-whiskers plots for responses to the *episode appropriateness* rating questions for experiment Group 2 (n =13).

Figure 46 shows box-and-whiskers plots for responses to the *episode appropriateness* rating questions to the four storyboards in Group 2. Thirteen participants viewed and answered questions about each of these four storyboards. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the response to the *episode appropriateness* rating question indicate that the responses to the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for the two control storyboards.

Mild outliers are labeled with open circles, as above. There are mild outliers for EA_26003 and EA_26011. In both cases, these outliers are counter to the hypothesized response pattern and were retained in the data set, on the grounds that retaining the outliers is more conservative for the analysis of the closed-response data. In the case of EA_26012, the responses labeled with participant IDs are marked by an asterisk to indicate that they are *extreme outliers*[46]. These data points were also retained because the status of these points as extreme outliers is simply a literal statistical truth that is a consequence of the fact that these are the only two responses to EA_26012 that differed from the mode response (which has a rating of 5). That is, all participants except those marked by the asterisk rated the appropriateness of the teacher's actions as *appropriate* for EA_26012. The two extreme outliers labeled the teacher's actions as *very appropriate.* The means and standard deviations for the *episode appropriateness* ratings are reported in Table 5

Table 5: Mean and standard deviation for the 4 *episode appropriateness* questions answered by Group 2 participants.

EA_(Storybaord) (*n*=13)	**Mean**	**Standard Deviation**
EA_26003	2.62	1.26
EA_26004	4.15	1.63
EA_26011	4.08	1.38
EA_26012	5.15	0.38

Table 5 reports the means and standard deviations for the *episode appropriateness* rating questions for Group 2. The EA(Group 2)$_{treatment}$ measure was derived from the means of EA_26003 and EA_26004. The EA(Group 2)$_{control}$ measure was derived from the means of EA_26011 and EA_26012. Results of significance testing for across-condition and within-group differences in means are reported below.

[46] *Extreme outliers* are data points beyond the outer fences (defined above).

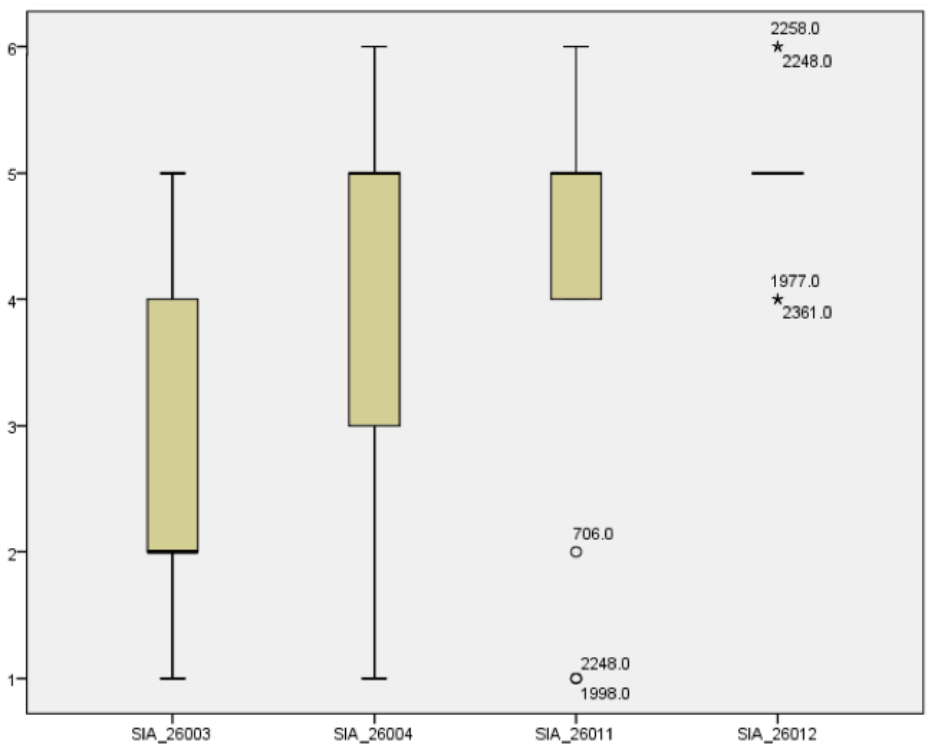

Figure 47: Box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions for experiment Group 2 (*n* =13).

Figure 47 shows box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions to the four storyboards in Group 2. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the responses to the *segment of interest* rating question indicate that the responses to each of the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for each of the control storyboards. As above, the responses marked by open circle are mild outliers; the responses indicated by the asterisks are extreme outliers—once again, these are responses that differed from the mode. The means and standard deviations for the *segment of interest appropriateness* ratings are reported in Table 6.

Table 6: Mean and standard deviation for the 4 *segment of interest appropriateness* questions answered by Group 2 participants.

$S_{(I)}A$**_(Storybaord) (*n*=13)**	**Mean**	**Standard Deviation**
$S_{(I)}A$ _26003	2.62	1.26
$S_{(I)}A$ _26004	4.00	1.63
$S_{(I)}A$ _26011	4.00	1.63
$S_{(I)}A$ _26012	5.00	0.58

Table 6 reports the means and standard deviations of the *segment of interest* rating questions for Group 2. The *segment of interest appropriateness* measures for Group 2 were derived from the means of corresponding storyboards. The $S_{(I)}A(\text{Group 2})_{treatment}$ measure was derived from the means of $S_{(I)}A$_26003 and $S_{(I)}A$_26004. The $S_{(I)}A\ (\text{Group 1})_{control}$ measure was derived from the means of $S_{(I)}A$_26011 and $S_{(I)}A$_26012. Results of significance testing for across-condition and within-group differences in means are reported below.

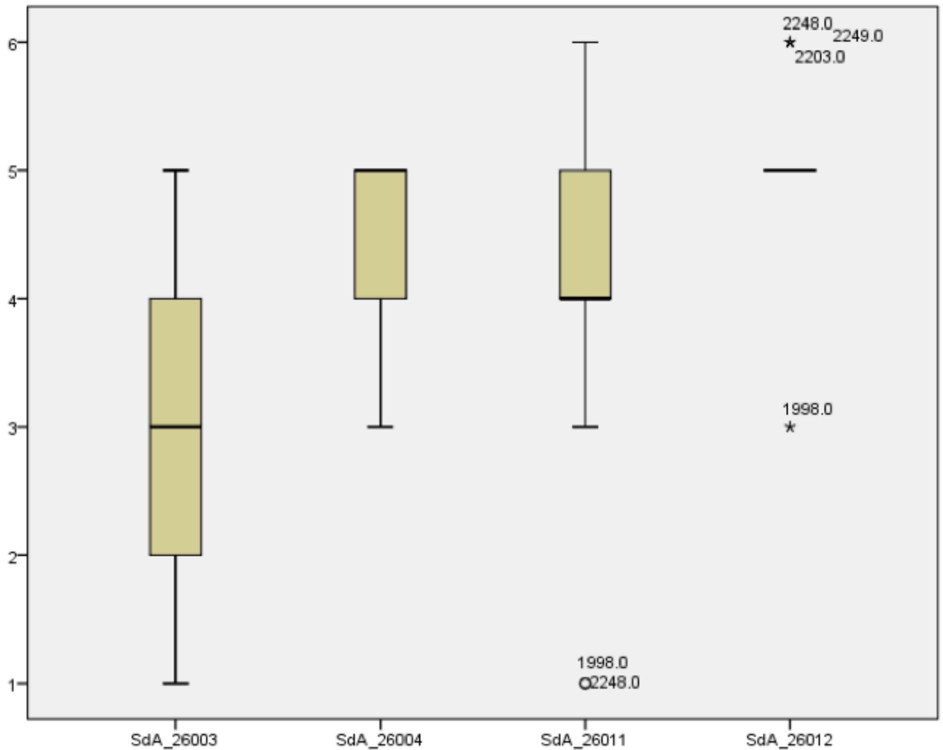

Figure 48: Box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions for experiment Group 2 (*n* =13).

Figure 48 shows box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions to the four storyboards in Group 1. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are marked as above. For three of the four Group 2 storyboards (i.e., all but 26003), the responses to the *distracter segment* rating questions indicate more positive reactions to the segment of the storyboard in which the teacher does an action that is unrelated to the norm. That responses to the *distracter segment* appropriates question for 26003 fit the same pattern as the responses to the *segment of interest* appropriateness question for 26003 could be an indication that the teacher's breach of the norm in the storyboard affected how participants rated consequent actions of the teacher. The means and standard deviations for the *distracter segment appropriateness* ratings are reported in Table 7.

Table 7: Mean and standard deviation for the 4 *distracter segment appropriateness* questions answered by Group 2 participants.

$S_{(D)}A$_(Storybaord) (*n*=13)	**Mean**	**Standard Deviation**
$S_{(D)}A$ _26003	3.23	1.42
$S_{(D)}A$ _26004	4.46	0.88
$S_{(D)}A$ _26011	4.00	1.53
$S_{(D)}A$ _26012	5.08	0.76

Table 7 shows the means and standard deviations of participants' responses to the *distracter segment* appropriateness rating questions in Group 2. The *distracter segment appropriateness* measures for Group 2 were derived from the means of corresponding storyboards. The $S_{(D)}A(\text{Group 2})_{treatment}$ measure was derived from the means of $S_{(D)}A$_26003 and $S_{(D)}A$_26004. The $S_{(D)}A(\text{Group\_2})_{control}$ measure was derived from the means of $S_{(D)}A$_26011 and $S_{(D)}A$_26012.

Descriptive statistics for experiment group 3.

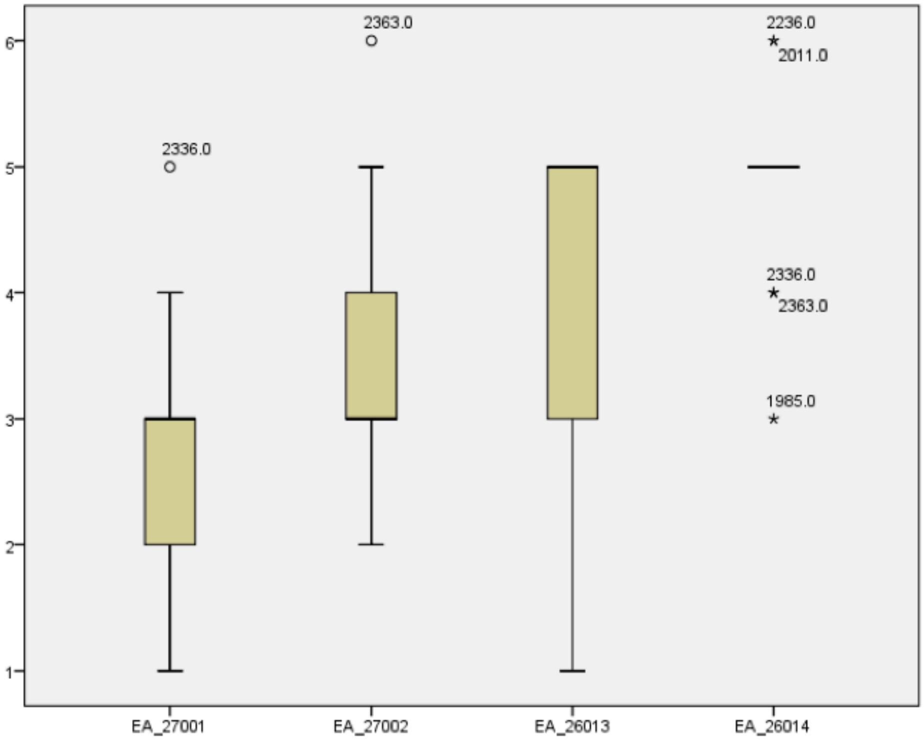

Figure 49: Box-and-whiskers plots for responses to the *episode appropriateness* rating questions for experiment Group 3 (n =15).

Figure 49 shows box-and-whiskers plots for responses to the *episode appropriateness* rating questions to the four storyboards in Group 3. Fifteen participants viewed and answered questions about each of these four storyboards. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are indicated as above. The plots of the responses to the *episode appropriateness* rating question indicate that the responses to the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for the two control storyboards. The means and standard deviations for the *episode appropriateness* ratings are reported in Table 8

Table 8: Mean and standard deviation for the 4 *episode appropriateness* questions answered by Group 3 participants.

EA_(Storybaord) (*n*=15)	**Mean**	**Standard Deviation**
EA_27001	2.67	1.11
EA_27002	3.40	1.12
EA_26013	4.00	1.31
EA_26014	4.15	1.63

Table 8 reports the means and standard deviations for the *episode appropriateness* rating questions for Group 3. The EA(Group_3)$_{treatment}$ measure was derived from the means of EA_27001 and EA_27002. The EA(Group 3)$_{control}$ measure was derived from the means of EA_27013 and EA_27014. Results of significance testing for across-condition and within-group differences in means are reported below.

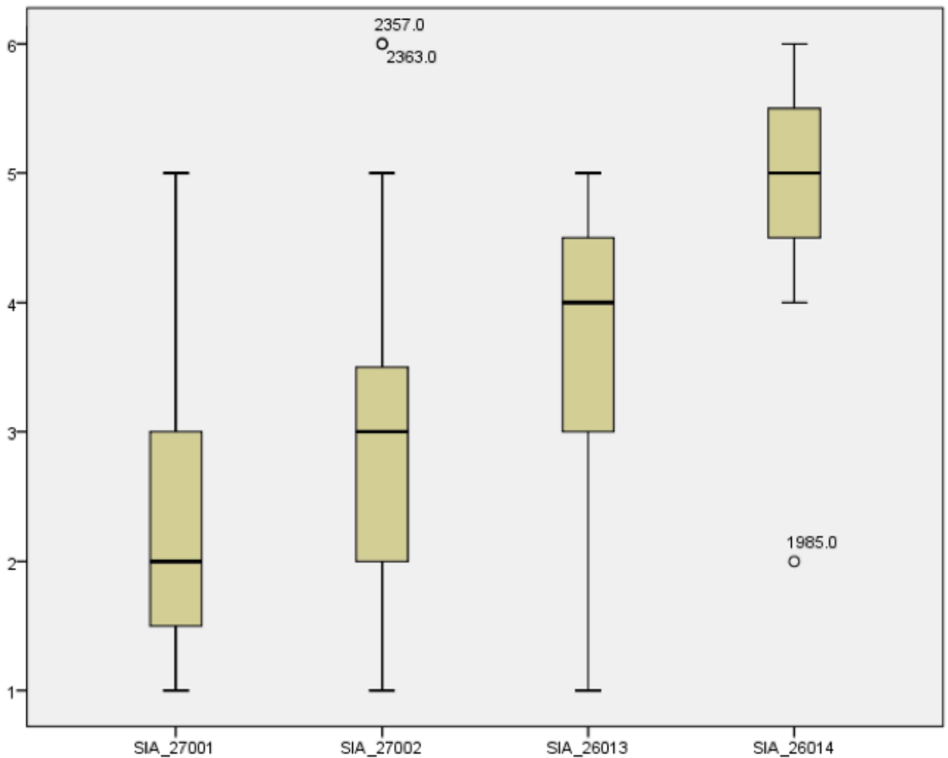

Figure 50: Box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions for experiment Group 3 (n =15).

Figure 50 shows box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions to the four storyboards in Group 3. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the responses to the *segment of interest* rating question indicate that the responses to each of the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for each of the control storyboards. As above, the responses marked by open circles are mild outliers. The means and standard deviations for the *segment of interest appropriateness* ratings are reported in Table 9.

Table 9: Mean and standard deviation for the 4 *segment of interest appropriateness* questions answered by Group 3 participants.

$S_{(I)}A$_**(Storybaord) (*n*=15)**	**Mean**	**Standard Deviation**
$S_{(I)}A$ _27001	2.47	1.36
$S_{(I)}A$ _27002	3.07	1.58
$S_{(I)}A$ _26013	3.60	1.18
$S_{(I)}A$ _26014	4.00	1.63

Table 9 reports the means and standard deviations for the *segment of interest* rating questions for Group 3. The *segment of interest appropriateness* measures for Group 3 were derived from the means of corresponding storyboards. The $S_{(I)}A(\text{Group 3})_{treatment}$ measure was derived from the means of $S_{(I)}A$_27001 and $S_{(I)}A$_27002. The $S_{(I)}A$ $(\text{Group 3})_{control}$ measure was derived from the means of $S_{(I)}A$_26013 and $S_{(I)}A$_26014. Results of significance testing for across-condition and within-group differences in means are reported below.

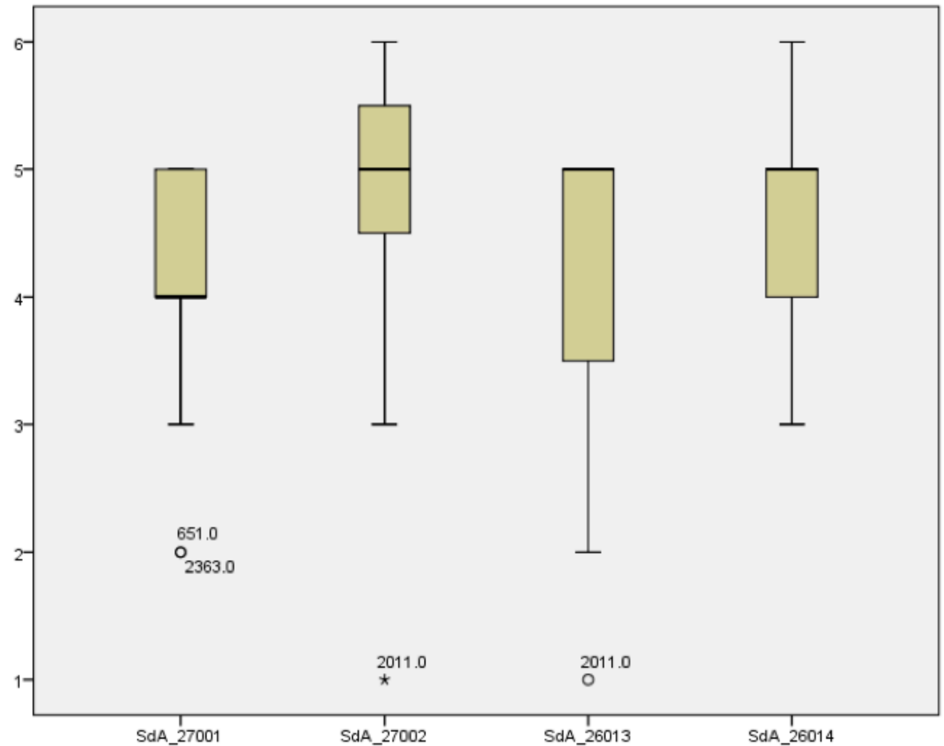

Figure 51: Box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions for experiment Group 3 (n =15).

Figure 51 shows box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions to the four storyboards in Group 3. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are marked as above. As was the case for all of the Group 1 storyboards and 3 of 4 Group 2 storyboards, the responses to the *distracter segment* rating questions indicate more balanced reactions to the segment of the storyboard in which the teacher does an action that is unrelated to the norm. The means and standard deviations for the *distracter segment appropriateness* ratings are reported in Table 10.

Table 10: Mean and standard deviation for the 4 *distracter segment appropriateness* questions answered by Group 3 participants.

$S_{(D)}$A_**(Storybaord) (*n*=15)**	**Mean**	**Standard Deviation**
$S_{(D)}$A _27001	4.07	1.03
$S_{(D)}$A _27002	4.73	1.33
$S_{(D)}$A _26013	4.13	1.41
$S_{(D)}$A _26014	4.46	0.88

Table 10 reports the means and standard deviations of participants' responses to the *distracter segment* appropriateness rating questions in Group 3. The *distracter segment appropriateness* measures for Group 3 were derived from the means of corresponding storyboards. The $S_{(D)}$A(Group 3)$_{treatment}$ measure was derived from the means of $S_{(D)}$A_27001 and $S_{(D)}$A_27002. The $S_{(D)}$A(Group_3)$_{control}$ measure was derived from the means of $S_{(D)}$A_26013 and $S_{(D)}$A_26014.

Descriptive statistics for experiment group 4.

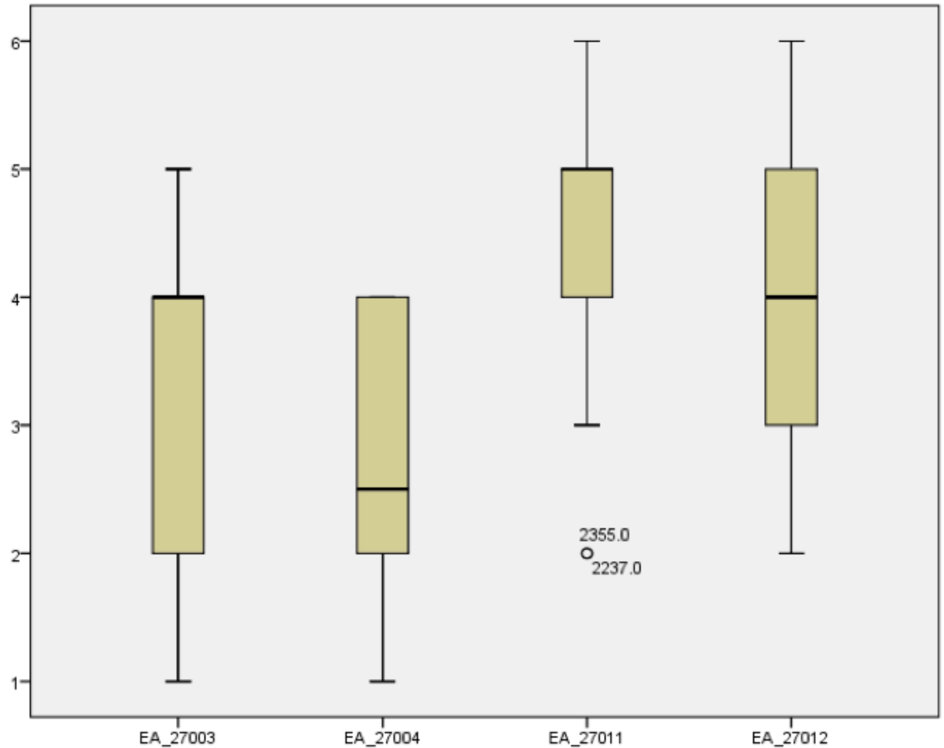

Figure 52: Box-and-whiskers plots for responses to the *episode appropriateness* rating questions for experiment Group 4 (*n* =14).

Figure 52 shows box-and-whiskers plots for responses to the *episode appropriateness* rating questions to the four storyboards in Group 4. Fourteen participants viewed and answered questions about each of these four storyboards. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are indicated as above. The plots of the responses to the *episode appropriateness* rating question indicate that the responses to the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for the two control storyboards. The means and standard deviations for the *episode appropriateness* ratings are reported in Table 11

Table 11: Mean and standard deviation for the 4 *episode appropriateness* questions answered by Group 4 participants.

EA_(Storybaord) (*n*=14)	**Mean**	**Standard Deviation**
EA_27003	3.36	1.34
EA_27004	2.79	1.05
EA_27011	4.57	1.34
EA_27012	4.07	1.27

Table 11 shows the means and standard deviations for the *episode appropriateness* rating questions for Group 4. The EA(Group_4)$_{treatment}$ measure was derived from the means of EA_27003 and EA_27004. The EA(Group 4)$_{control}$ measure was derived from the means of EA_27011 and EA_27012. Results of significance testing for across-condition and within-group differences in means are reported below.

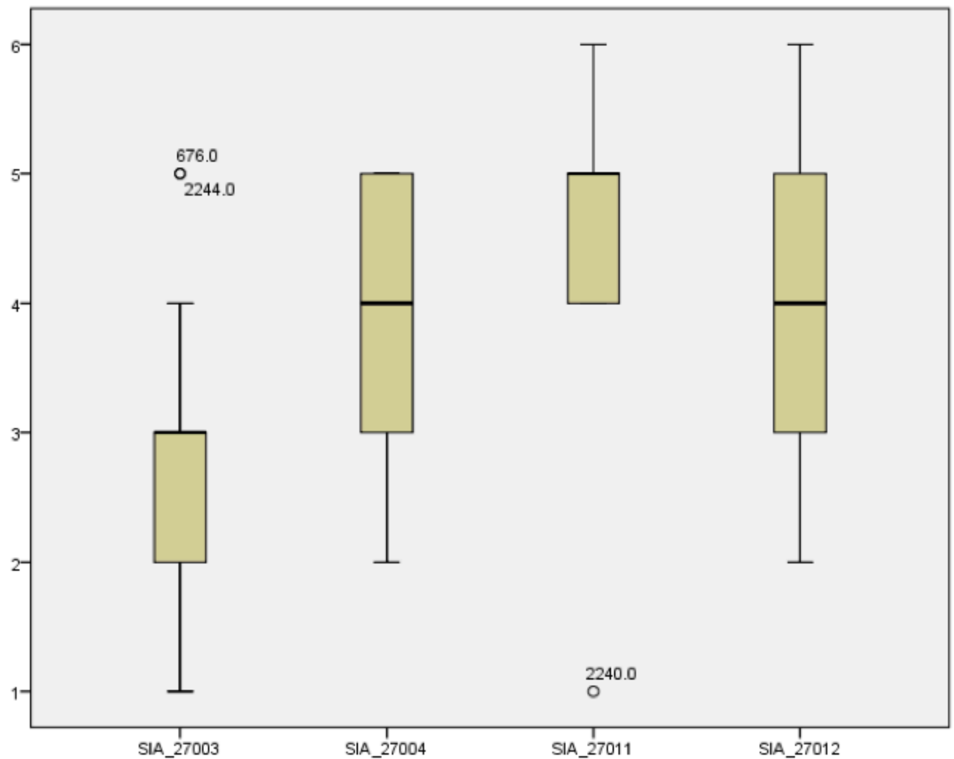

Figure 53: Box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions for experiment Group 4 (n =14).

Figure 53 shows box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions to the four storyboards in Group 4. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the responses to the *segment of interest* rating question indicate that the responses to each of the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for each of the control storyboards. As above, the responses marked by open circles are mild outliers. The means and standard deviations for the *segment of interest appropriateness* ratings are reported in Table 12.

Table 12: Mean and standard deviation for the 4 *segment of interest appropriateness* questions answered by Group 4 participants.

$S_{(I)}A$**_(Storybaord) (*n*=14)**	**Mean**	**Standard Deviation**
$S_{(I)}A$ _27003	2.79	1.25
$S_{(I)}A$ _27004	4.00	0.96
$S_{(I)}A$ _27011	4.64	1.28
$S_{(I)}A$ _27012	3.86	1.41

Table 12 shows the means and standard deviations for the *segment of interest* rating questions for Group 4. The *segment of interest appropriateness* measures for Group 4 were derived from the means of the corresponding storyboards. The $S_{(I)}A(\text{Group 4})_{treatment}$ measure was derived from the means of $S_{(I)}A$_27003 and $S_{(I)}A$_27004. The $S_{(I)}A$ $(\text{Group 1})_{control}$ measure was derived from the means of $S_{(I)}A$_27011 and $S_{(I)}A$_27012. Results of significance testing for across-condition and within-group differences in means are reported below.

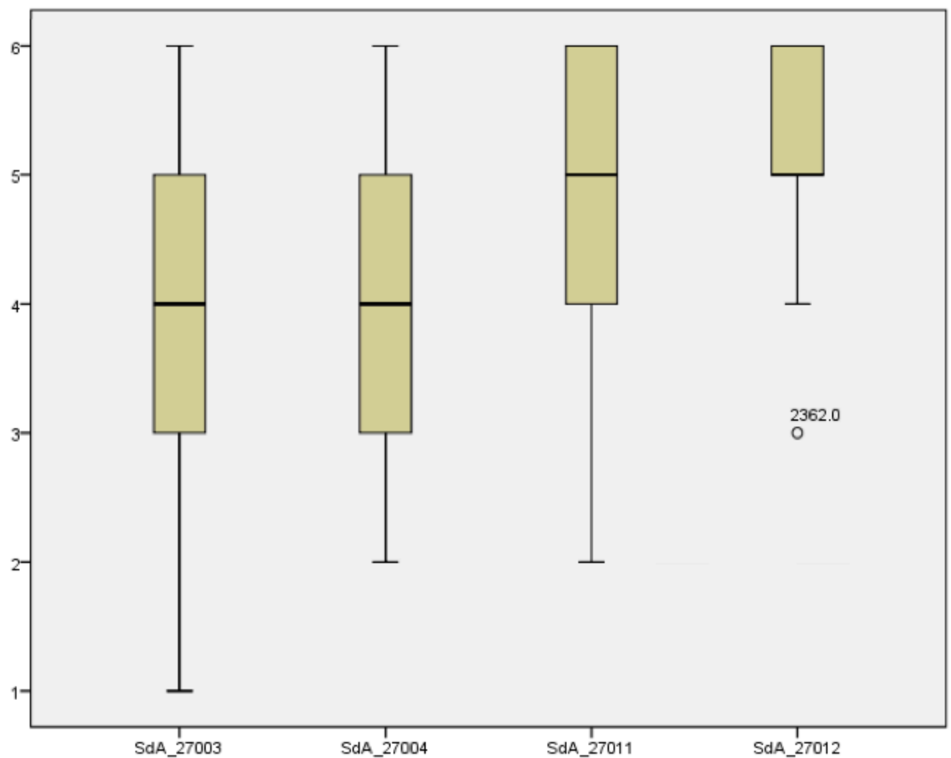

Figure 54: Box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions for experiment Group 4 (n =14).

Figure 54 shows box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions to the four storyboards in Group 4. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are marked as above. The means and standard deviations for the *distracter segment appropriateness* ratings are reported in Table 13.

Table 13: Mean and standard deviation for the 4 *distracter segment appropriateness* questions answered by Group 4 participants.

$S_{(D)}$A_**(Storybaord) (*n*=14)**	**Mean**	**Standard Deviation**
$S_{(D)}$A _27003	3.86	1.29
$S_{(D)}$A _27004	3.93	1.07
$S_{(D)}$A _27011	4.64	1.22
$S_{(D)}$A _27012	5.14	0.86

Table 13 shows the means and standard deviations of participants' responses to the *distracter segment* appropriateness rating question. The *distracter segment appropriateness* measures for Group 4 were derived from the means of corresponding storyboards. The $S_{(D)}A(Group\ 4)_{treatment}$ measure was derived from the means of $S_{(D)}A\_27003$ and $S_{(D)}A\_27004$. The $S_{(D)}A(Group\_4)_{control}$ measure was derived from the means of $S_{(D)}A\_27011$ and $S_{(D)}A\_27012$.

Descriptive statistics for experiment group 5.

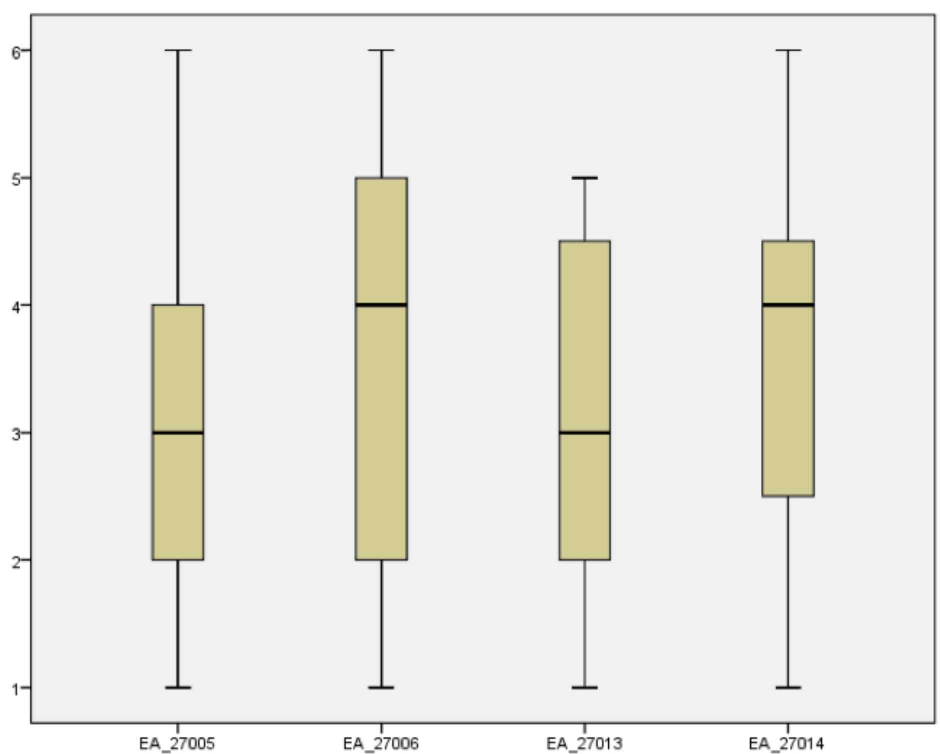

Figure 55: Box-and-whiskers plots for responses to the *episode appropriateness* rating questions for experiment Group 5 (n =15).

Figure 55 shows box-and-whiskers plots for responses to the *episode appropriateness* rating questions to the four storyboards in Group 5. Fifteen participants viewed and answered questions about each of these four storyboards. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the responses to the *episode appropriateness* rating question for Group 5 break with the general pattern for responses to the *episode appropriateness* ratings questions that were described above. For Group 5, there median and quartile scores for the breach storyboards are generally greater than or equal to the median and quartile scores for the control storyboards. Also, compared to the previous *episode appropriateness* descriptive statistics reported above, the control storyboards yielded ratings that tended to be lower. This issue is investigated during the reporting of the hypothesis testing. The means and standard deviations for the *episode appropriateness* ratings are reported in Table 14.

Table 14: Mean and standard deviation for the 4 *episode appropriateness* questions answered by Group 5 participants.

EA_(Storybaord) (*n*=14)	**Mean**	**Standard Deviation**
EA_27005	3.13	1.55
EA_27006	3.87	1.73
EA_27013	3.20	1.47
EA_27014	3.67	1.50

Table 14 reports the means and standard deviations of *episode appropriateness* questions for Group 5. The breach storyboards produced the highest (3.87) and lowest (3.13) means in the set. How the breach and control storyboards in Group 5 functioned in comparison to the storyboards in the other groups is considered below. The results reported in Table 14 and in the box-and-whiskers representation of the data for Group 5 suggests that there were not clear differences between the appropriateness of the teacher's actions in the breach and control storyboards. The means of the breach storyboards in Group 5 are in the neighborhood of the means for the breach storyboards in Groups 1-4; the difference is in the means of the control storyboards. These are lower than the means for the control storyboards on other groups. I explore this issue further during the significance testing and in the discussion of the results.

The EA(Group 5)$_{treatment}$ measure was derived from the means of EA_27005 and EA_27006. The EA(Group 5)$_{control}$ measure was derived from the means of EA_27013 and EA_27014. Results of significance testing for across-condition and within-group differences in means are reported below.

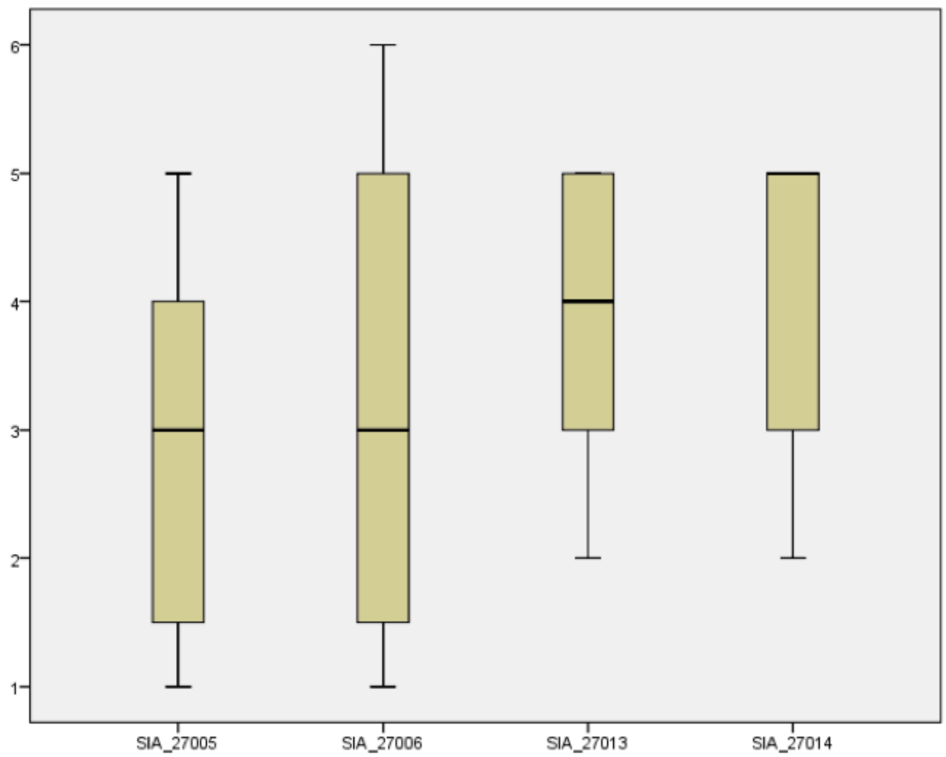

Figure 56: Box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions for experiment Group 5 (n =15).

Figure 56 shows box-and-whiskers plots for responses to the *segment of interest appropriateness* rating questions to the four storyboards in Group 5. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. The plots of the responses to the *segment of interest* rating question indicate that the responses to each of the breach storyboards had median, first-quartile, and third-quartile scores that were less than or equal to the median, first-quartile, and third-quartile scores for each of the control storyboards. However, as above, the ratings for the *segment of interest* appropriateness for the control storyboards in Group 5 appear to tend toward more negative ratings of the teacher. The means and standard deviations for the *segment of interest appropriateness* ratings are reported in Table 15.

Table 15: Mean and standard deviation for the 4 *segment of interest appropriateness* questions answered by Group 5 participants.

$S_{(I)}A$**_(Storybaord) (*n*=14)**	**Mean**	**Standard Deviation**
$S_{(I)}A$ _27005	2.80	1.52
$S_{(I)}A$ _27006	3.07	1.87
$S_{(I)}A$ _27013	3.87	1.19
$S_{(I)}A$ _27014	4.00	1.25

Table 15 reports the means and standard deviations of the *segment of interest* rating questions for Group 5. The *segment of interest appropriateness* measures for Group 5 were derived from the means of the corresponding storyboards. The $S_{(I)}A(\text{Group 5})_{treatment}$ measure was derived from the means of $S_{(I)}A$_27005 and $S_{(I)}A$_27006. The $S_{(I)}A$ $(\text{Group 1})_{control}$ measure was derived from the means of $S_{(I)}A$_27013 and $S_{(I)}A$_27014. Results of significance testing for across-condition and within-group differences in means are reported below.

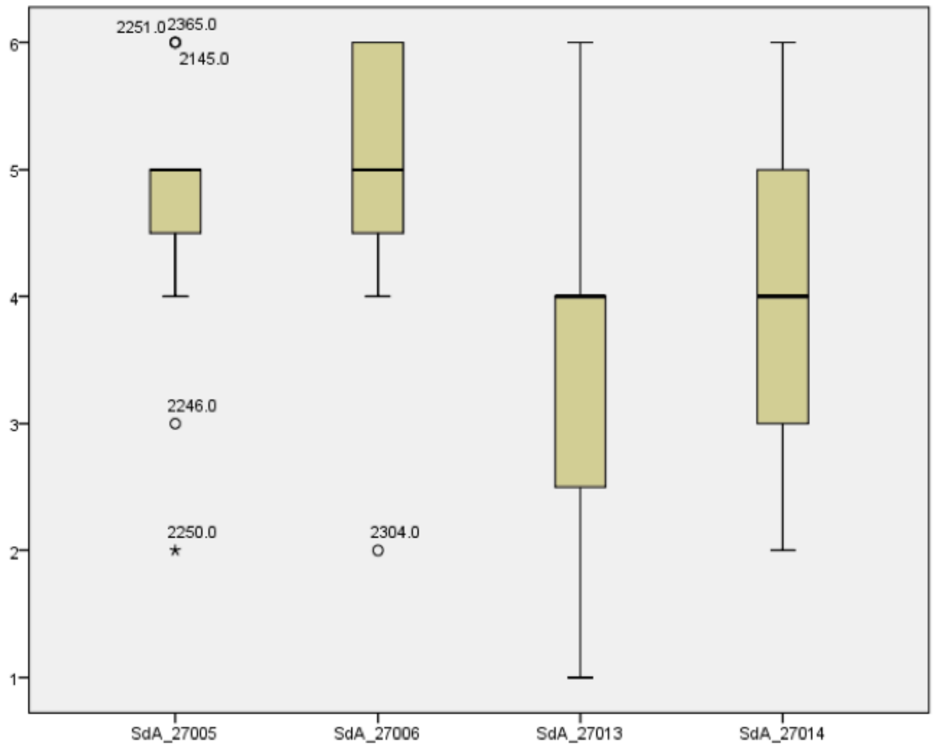

Figure 57: Box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions for experiment Group 5 (n =15).

Figure 57 shows box-and-whiskers plots for responses to the *distracter segment appropriateness* rating questions to the four storyboards in Group 5. The breach storyboards are the two leftmost box-and-whiskers plots. The control storyboards are the two rightmost box-and-whiskers plots. Outliers are marked as above. The responses to the two control storyboards have lower medians and lower inner-fence boundaries than the breach storyboards. I revisit this issue during the report of the hypothesis testing. The means and standard deviations for the *distracter segment appropriateness* ratings are reported in Table 16.

Table 16: Mean and standard deviation for the 4 *distracter segment appropriateness* questions answered by Group 5 participants.

$S_{(D)}$A_(Storybaord) (*n*=15)	Mean	Standard Deviation
$S_{(D)}$A _27005	4.73	1.10
$S_{(D)}$A _27006	5.00	1.13
$S_{(D)}$A _27013	3.53	1.36
$S_{(D)}$A _27014	4.00	1.20

Table 16 shows the means and standard deviations of participants' responses to the *distracter segment* appropriateness rating questions in Group 5. The *distracter segment appropriateness* measures for Group 5 were derived from the means of corresponding storyboards. The $S_{(D)}$A(Group 5)$_{treatment}$ measure was derived from the means of $S_{(D)}$A_27005 and $S_{(D)}$A_27006. The $S_{(D)}$A(Group_5)$_{control}$ measure was derived from the means of $S_{(D)}$A_27013 and $S_{(D)}$A_27014.

Summary of descriptive statistics

In this section, I presented box-and-whiskers plots of the closed-response data participants provided after viewing each storyboard. The results were reported by experiment group and question type. I defined the EA, $S_{(I)}$A, and $S_{(D)}$A measures in terms of corresponding storyboards from each experiment group. Below, I use these measures to state and test hypotheses about participants' rating responses to the different storyboards.

Hypothesized differences between measures

The general hypothesis underlying the design of the instrument was that participants would react negatively to a teacher breaching a hypothesized semiotic norm. The treatment/control design using matched storyboards makes it possible to define "negative reactions" in terms of differences in the EA, $S_{(I)}A$, and $S_{(D)}A$ measures defined above. Since a norm is not only what is routine but also what is expected, I hypothesized that participants would find the work of the teacher less appropriate in storyboards that breached a hypothetical norm when compared to storyboards where a norm was not breached. This general hypothesis comprises 5 specific hypotheses that can be stated in terms of the means of the episode and segment appropriateness measures.

Across-Group Hypotheses:

I. H_0: $EA_{treatment} = EA_{control}$
H_1: $EA_{treatment} < EA_{control}$

II. H_0: $S_{(I)}A_{treatment} = S_{(I)}A_{control}$
H_1: $S_{(I)}A_{treatment} < S_{(I)}A_{control}$

III. H_0: $S_{(D)}A_{treatment} = S_{(D)}A_{control}$
H_1: $S_{(D)}A_{treatment} \neq S_{(D)}A_{control}$

Within-Group Hypotheses:

IV. H_0: $S_{(I)}A_{treatment} = S_{(D)}A_{treatment}$
H_1: $S_{(I)}A_{treatment} < S_{(D)}A_{treatment}$

V. H_0: $S_{(I)}A_{control} = S_{(D)}A_{control}$
H_1: $S_{(I)}A_{control} \neq S_{(D)}A_{control}$

There are three across-group hypotheses (I-III) and two within-group hypotheses (IV, V). The three-across group hypotheses are treatment/control comparisons between participants that viewed different versions of a storyboard. For these comparisons, for the same target norm, the mean appropriateness ratings to questions following storyboards where the norm is breached were compared to mean appropriateness ratings to questions following storyboards in which the norm is not breached. The two within-group hypotheses test for differences in how participants rated segments of the episodes where the teacher is shown as breaching (or not breaching) a norm and some other segment. The five hypotheses were tested for each of the five norms, yielding five groups of five hypothesis tests. The results of these tests are reported below.

The statements of the hypotheses shown above are generic statements of the 5 comparisons that were made for each norm that was investigated during the experiment. I tested for significant differences using independent (I, II, III) and paired-samples (IV, V) *t*-tests because of the matched design of the planned comparisons. Norm-specific hypotheses were tested for each matched pair of EA, $S_{(I)}A$, and $S_{(D)}A$ measures. For example, the *less details* norm used Group 1 as the treatment and Group 2 as the control. Thus, the mean of $EA(Group1)_{treatment}$ was compared to the mean of $EA(Group2)_{control}$ to test Hypothesis I for the *less details* norm.

Alternate hypotheses (I), (II) and (IV) are directed because I predicted the treatment storyboards will have lower mean ratings than the control storyboards. Alternate hypotheses (III) and (V) are non-directed, because I predicted that there will not be significant differences between mean ratings for $S_{(D)}A$. That is, for hypotheses (III) and (V), I predicted that the null hypothesis would *not* be rejected. All hypotheses were tested for significance against a two-tailed distribution. The next section reports the norm-by-norm results of the significance tests.

Results of Significance Testing

Details norm hypothesis testing

The two treatment storyboards in experiment Group 1 were episodes in which the teacher was represented as breaching the normative ways that the details of a proof are checked. The treatment storyboards in Group 1 were matched with control storyboards that were administered to participants in Group 2. For the treatment storyboards, the teacher enacts the breach by accepting a proof that omits a detail that is hypothesized to be required. These are the *less details* realizations of breaches to the *details* norm. For the control storyboards, the teacher insists that these details be included before the proof can be accepted. The teacher's action in the control storyboards is hypothesized to comply with the details norm. Table 17 and Table 18 report the results of the significance testing of hypotheses I-V for the Group 1 and Group 2 episode and segment appropriateness measures.

Table 17: Results of across-group hypothesis testing (I, II, III) comparing means of Group 1 (treatment) measures with means of Group 2 (control) measures.

Measure	**Treatment (*n* = 16)**	**Control (*n*=13)**	μ_1-μ_2	*p*
EA (I)	3.469	4.615	-1.146	.001*
$S_{(I)}A$ (II)	2.312	4.5	-2.187	<.001*
$S_{(D)}A$ (III)	4.219	4.538	-0.319	.374

*Significant at the .05 level.

Table 18: Results of within-group hypothesis testing (IV, V) comparing means of *segment appropriateness measures* for treatment (Group 1, *n* =16) and control (Group 2, *n*=13)..

Condition	**$S_{(I)}A$**	**$S_{(D)}A$**	μ_1-μ_2	*p*
Treatment (IV)	2.312	4.219	-1.146	<.001*
Control (V)	4.781	4.875	-0.094	.704

*Significant at the .05 level.

The results reported in Table 17 and Table 18 show that the observed differences between the mean ratings for the EA, $S_{(I)}A$, and $S_{(D)}A$ measures were as predicted. The means on the episode appropriateness rating questions were significantly lower for the storyboards where the teacher was depicted a breaching the norm when compared to those storyboards where the teacher did not breach a norm.

The two treatment storyboards in experiment Group 2 were episodes in which the teacher was also represented as breaching the normative ways that the details of a proof are checked. In the case of Group 2, however, the teacher breaches the details norm by insisting that the students include details in their proofs that are hypothesized to be unnecessary, because they are details that are conveyed by the diagram that accompanies the proof. For the control storyboards, the teacher does not insist that the proofs include such detail. The treatment storyboards in Group 2 were matched with control storyboards that were administered to participants in Group 3. The treatment storyboards in Group 2 are the *more details* realizations of breaches to the *details* norm. Table 19 and Table 20 report the results of the significance testing of hypotheses I-V for the Group 2 and Group 3 episode and segment appropriateness measures.

Table 19: Results of across-group hypothesis testing (I, II, III) comparing means of Group 2 treatment measures with means of Group 3 control measures.

Measure	**Treatment (n = 13)**	**Control (n=15)**	μ_1-μ_2	p
EA (I)	3.385	4.433	-1.049	.006*
$S_{(I)}$A (II)	3.307	4.233	-.926	.009*
$S_{(D)}$A (III)	3.846	4.366	-0.520	.115

*Significant at the .05 level.

Table 20: Results of within-group hypothesis testing (IV, V) comparing means of *segment appropriateness measures* for treatment (Group 2, n =13) and control (Group 3, n =15).

Condition	**$S_{(I)}$A**	**$S_{(D)}$A**	μ_1-μ_2	p
Treatment (IV)	3.307	3.846	-.538	.024*
Control (V)	4.233	4.366	-0.267	.499

*Significant at the .05 level.

The results reported in Table 19 and Table 20 show that, for the across-group comparisons for the *more details* norm, the observed differences between the mean ratings for the EA, $S_{(I)}$A, and $S_{(D)}$A measures were as predicted. The means on the episode appropriateness rating questions were significantly lower for the storyboards where the teacher was depicted breaching the norm when compared to those storyboards where the teacher did not breach a norm. For the within group comparisons for Group 2, there was no significant difference observed between the segments of interest measures for the control storyboards. This was as predicted. Also as predicted, in the storyboard that depicted a breach of details norm, the mean for the segment that targeted the breach was significantly lower than the mean for the segment that targeted the common ground.

Sequence norm hypothesis testing

The two treatment storyboards in experiment Group 3 were episodes in which the teacher was represented as breaching the labels aspect of the sequence norm. The sequence norm is based on the the hypothesis that when students are called to the board to present proofs in geometry, the task for the student is primarily that of transcription: The student is expected to reproduce a proof that has been written out already (e.g., on the student's homework), mark-for-mark on the board. The labels aspect of the sequence norm focuses on the order in which the labels of a diagram are reproduced during this transcription.

For the treatment storyboards in Group 3, the teacher breaches the sequence norm by taking issue with the order in which the student presenting the proof reproduces the labels on the diagram. The teacher insists that the student write the labels as they are needed in the development of the proof, rather than allow the student to add the labels when it is convenient for the student to do so (as would be expected if the teacher were complying with the hypothesized norm). The treatment storyboards in Group 3 were matched with control storyboards that were administered to participants in Group 4. The treatment storyboards in Group 3 are the *sequence (labels)* realizations of breaches to the *sequence* norm. Table 21 and Table 22 report the results of the significance testing of hypotheses I-V for the Group 3 and Group 4 episode and segment appropriateness measures.

Table 21: Results of across-group hypothesis testing (I, II, III) comparing means of Group 3 treatment measures with means of Group 4 control measures.

Measure	Treatment (*n* = 15)	Control (*n*=14)	μ_1-μ_2	*p*
EA (I)	3.305	4.432	-1.049	.002*
$S_{(I)}$A (II)	2.766	4.25	-.926	.001*
$S_{(D)}$A (III)	4.4	4.892	-0.520	.059

*Significant at the .05 level.

Table 22: Results of within-group hypothesis testing (IV, V) comparing means of *segment appropriateness measures* for treatment (Group 3, *n* =15) and control (Group 4, *n* = 14).

Condition	$S_{(I)}$A	$S_{(D)}$A	μ_1-μ_2	*p*
Treatment (IV)	2.766	4.4	-1.633	<.001*
Control (V)	4.25	4.893	-0.643	.089

*Significant at the .05 level.

The results reported in Table 21 and Table 22 show that, for the across-group comparisons for the *sequence (labels)* norm, the observed differences between the mean ratings for the EA, $S_{(I)}$A, and $S_{(D)}$A measures were as predicted. The means on the episode appropriateness rating questions were significantly lower for the storyboards where the teacher was depicted breaching the norm when compared to those storyboards where the teacher did not breach a norm. Furthermore, while there was a difference observed between the measures for the distracter segment of interest, this difference was not significant. The observed within-group differences were also as predicted: Participants rated lower the work of the teacher in the segments of the storyboards where the teacher was shown breaching the sequence norm by insisting that the student label the diagram before using the labels in the proof (Hypothesis IV). It was also the case that the difference in mean ratings between the segments of interest for the control storyboards in Group 4 were not significant, as predicted.

The two treatment storyboards in experiment Group 4 were episodes in which the teacher was represented as breaching the markings aspect of the sequence norm. The markings aspect of the sequence norm focuses on the order in which the markings of a diagram are reproduced during the transcription of a proof. For the treatment storyboards in Group 4, the teacher breaches the sequence norm by taking issue with the order in which the student presenting the proof reproduces the markings on the diagram. The teacher insists that the student mark the diagram as the properties indicated by the diagram become established in the proof, rather than allow the student to add the markings when it is convenient for the student to do so (as would be expected if the teacher were complying with the hypothesized norm).

The treatment storyboards in Group 4 were matched with control storyboards that were administered to participants in Group 5. The treatment storyboards in Group 4 are the *sequence (markings)* realizations of breaches to the *sequence* norm. Table 23 and Table 24 report the results of the significance testing of hypotheses I-V for the Group 4 and Group 5 episode and segment appropriateness measures.

Table 23: Results of across-group hypothesis testing (I, II, III) comparing means of Group 4 treatment measures with means of Group 5 control measures.

Measure	**Treatment (n = 14)**	**Control (n=15)**	μ_1-μ_2	p
EA (I)	3.071	3.433	-.361	.374
$S_{(I)}$A (II)	3.392	3.933	-.540	.119
$S_{(D)}$A (III)	3.892	3.766	.126	.722

Table 24: Results of within-group hypothesis testing (IV, V) comparing means of *segment appropriateness measures* for treatment (Group 4, n =14) and control (Group 5, n = 15).

Condition	**$S_{(I)}A$**	**$S_{(D)}A$**	μ_1-μ_2	p
Treatment (IV)	3.393	3.893	-.5	.089
Control (V)	3.857	3.893	-0.072	.036*

*Significant at the .05 level.

The results reported in Table 23 and Table 24 show that none of the observed differences for ratings on the episode or segment of interest measures were significant. This was as predicted for Hypotheses III and V, however for the other hypotheses, the means of the measures derived from the ratings questions that corresponded to the treatment storyboards were not significantly lower than the means of the measures that were derived from the ratings questions that corresponded to the control storyboards. The fact that the observed differences did not match the predictions suggests that the markings aspect of the sequence norm was not recognized by the participants in the study. I revisit this issue during the discussion of the quantitative findings and the report of the analysis of open response data in Chapter 6.

The two treatment storyboards in experiment Group 5 were episodes in which the teacher was represented as breaching the reasons aspect of the sequence norm. The reasons aspect of the sequence norm focuses on the order in which the reasons of a proof are reproduced during the transcription of a proof. For the treatment storyboards in Group 5, the teacher breaches the sequence norm by taking issue with the order in which the student presenting the proof reproduces the reasons of the proof. The teacher insists that the student include the reasons as they become necessary to warrant the statements that are established in the proof, rather than allow the student to add the reasons when it is convenient for the student to do so (as would be expected if the teacher were complying with the hypothesized norm).

The treatment storyboards in Group 5 were matched with control storyboards that were administered to participants in Group 1. The treatment storyboards in Group 5 are the *sequence (reasons)* realizations of breaches to the *sequence* norm. Table 25 and Table 26 report the results of the significance testing of hypotheses I-V for the Group 5 and Group 1 episode and segment appropriateness measures.

Table 25: Results of across-group hypothesis testing (I, II, III) comparing means of Group 5 treatment measures with means of Group 1 control measures.

Measure	**Treatment (n = 15)**	**Control (n=16)**	μ_1-μ_2	p
EA (I)	3.5	4.531	-1.031	.023*
$S_{(I)}A$ (II)	2.933	4.781	-1.847	<.001*
$S_{(D)}A$ (III)	4.866	4.875	-.008	.981

*Significant at the .05 level.

Table 26: Results of within-group hypothesis testing (IV, V) comparing means of *segment appropriateness measures* for treatment (Group 5, n =15) and control (Group 1, n = 16).

Condition	$\mathbf{S_{(I)}A}$	$\mathbf{S_{(D)}A}$	μ_1-μ_2	p
Treatment (IV)	2.933	4.866	-1.933	<.001
Control (V)	4.781	4.875	-0.093	.704

*Significant at the .05 level.

The results reported in Table 25 and Table 26 show that, for the across-group comparisons for the *sequence (reasons)* norm, the observed differences between the mean ratings for the EA, $S_{(I)}A$, and $S_{(D)}A$ measures were as predicted. The means on the episode appropriateness rating questions were significantly lower for the storyboards where the teacher was depicted breaching the norm when compared to those storyboards where the teacher did not breach a norm. Furthermore, while there was a difference observed between the measures for the distracter segment of interest, this difference was not significant. The observed within-group differences were also as predicted: Participants rated lower the work of the teacher in the segments of the storyboards where the teacher was shown breaching the sequence norm by requiring the student to include the reasons as they were logically used in the argument, rather than in the order in which the student found it convenient to transcribe those reasons when copying the proof on the board (Hypothesis IV). It was also the case that the difference in mean ratings between the segments of interest (Hypothesis V) for the control storyboards in Group 1 were not significant, as predicted.

The results reported above show that the observed differences between episode and segment appropriateness measures were, for the most part, as predicted. The exception to this general statement is the results of the analysis for the markings aspect of the sequence norm. The next section considers the validity of the EA, $S_{(I)}A$, and $S_{(D)}A$ measures.

Assessing the validity of the EA, $S_{(I)}A$, and $S_{(D)}A$ measures

By design, each participant in the study viewed and answered questions about two storyboards that breached a target semiotic norm and two other storyboards. The two other storyboards were the control duals for an unrelated semiotic norm. The purpose of having each participant view storyboards in pairs (breach pair, control pair) was to be able to derive *episode* and *segment appropriateness* measures that were not directly tied to the particulars of single storyboards. A basic indicator of the reliability of these measures is the correlation between responses to corresponding questions on the breach and control storyboard pairs in each experiment group. For example, the *episode appropriateness* ratings for the two breach storyboards in an experiment were expected to be correlated, as were the *episode appropriateness* ratings for the two control storyboards within each experiment group. Likewise for the *segment of interest* and *distracter segment* ratings for the two breach and two control storyboards in each experiment group.

In general, responses to corresponding questions on breach and control storyboards did not correlate within the experiment groups. This is the case for Pearson's product-moment correlation as well as for Spearman's Rho and Kendall's Tau. That the scores do not, in general, correlate is somewhat surprising but not inexplicable. The measures that were derived from the responses to the various appropriateness rating questions are functionally Likert-like scales that consist of two items each (Bonett, 2002; Carifio & Perla, 2007; Cohen, Manion & Morrison, 2007 p. 506). By condition (breach/control), every participant was measured using three scales: the Episode Appropriateness scale, the Segment of Interest Appropriateness scale, and the Distracter Segment Appropriateness scale. The Pearson correlation has been shown to be an inadequate measure of reliability for two-item scales (Eisinga et al., 2013). Furthermore, reliability tends to decrease as the number of items and sample size decrease (Bonett, 2002). The study conducted for this dissertation faced the practical challenge of a small number of participants (per experiment group) taking a small number of items: fewer than 20 participants for each condition and two items per scale. The fact that the items did not correlate is not necessarily an indictment of the items, but rather could be a product of the small sample sizes and low numbers of items. From this perspective, the lack of significant correlations between corresponding items in an experiment group is inconclusive.

For the data collection for this study, I investigated a larger set of hypothetical semiotic norms at the expense of having fewer participants available in each experiment group. The analysis reported above provides initial indications that the experiment worked as it was designed and that secondary teachers recognize 4 of the 5 semiotic norms that were tested during the study. These initial findings warrant additional data collection that could prioritize building a critical mass of participants to power a reliability analysis.

To increase the number of items that would be used to define the various measures, a future data collection could stratify experiment groups by condition. For example, one group of participants could view all four storyboards that represent breaches of the *details* norm, and a corresponding group could view the four control duals of these storyboards. This would double the number of items that would make up the EA, $S_{(I)}A$, and $S_{(D)}A$ measures for the details norm.

Even without collecting additional data, the fact that responses to corresponding questions do not correlate does not mean it is unwarranted to define the measures I described above. Another indicator of the validity of the measures is the results of storyboard-by-storyboard hypothesis testing. A supplementary analysis that tests hypotheses I-V at a storyboard-by-storyboard level showed general consistency with the results reported above. The consistency of the results at the storyboard-by-storyboard level provides an indicator that the effects reported above are not artefacts of arbitrary groupings of storyboards, but rather capture participants' general tendencies to rate lower the actions of the teacher in storyboards (episode appropriateness) and specific segments of storyboards (segment of interest appropriateness) in which a teacher is shown departing from a hypothesized semiotic norm. Beyond helping to warrant the measures that were derived from the closed-response data, the storyboard-by-storyboard testing also provides an opportunity to investigate why the hypothesis tests that compared experiment Group 4 to experiment Group 5 were contrary to what was predicted.

Storyboard-by-Storyboard hypothesis testing

The hypotheses that are stated above in terms of the EA, $S_{(I)}A$, and $S_{(D)}A$ measures were tested storyboard-by-storyboard. For these comparisons, mean responses to the *episode appropriateness* and *segment appropriateness* rating questions were compared across breach/control storyboard pairs. Responses to the *segment appropriateness* rating questions were also compared within storyboards. For example, EA, $S_{(I)}A$, and $S_{(D)}A$ responses to rating questions following storyboard 26001 were compared to EA, $S_{(I)}A$, and $S_{(D)}A$ responses to rating questions following storyboard 26011. In addition, responses to $S_{(I)}A$ for storyboard 26001 were compared to responses to $S_{(D)}A$ for storyboard 26001 (likewise for 26011). The results of the storyboard-by-storyboard significance testing are reported hypothesis-by-hypothesis.

Across-groups hypothesis testing

In terms of storyboard comparisons, the across-group hypotheses are:

Across-Group Hypotheses:

I. H_0: EA(breach storyboard) = EA(control storyboard)
H_1: EA(breach storyboard) < EA(control storyboard)

II. H_0: $S_{(I)}A$(breach storyboard) = $S_{(I)}A$(control storyboard)
H_1: $S_{(I)}A$(breach storyboard)< $S_{(I)}A$(control storyboard)

III. H_0: $S_{(D)}A$(breach storyboard) = $S_{(D)}A$(control storyboard)
H_1: $S_{(D)}A$(breach storyboard) ≠ $S_{(D)}A$(control storyboard)

The above hypotheses are stated in terms of generic storyboards. For each breach/control storyboard pair (i.e., a breach storyboard and its control dual), hypotheses I-III were tested using independent samples *t*-tests. The results of these significance tests are reported by hypothesis and storyboard pair in Table 27-Table 29 below.

Table 27: Results of significance testing for Hypothesis I: EA(breach storyboard) < EA(control storyboard)

Storyboard pair (breach/control)	*N* (breach)	*N* (control)	μ_1 (breach)	μ_2 (control)	μ_1-μ_2	*p*
26001/11	16	13	3.313	4.077	-0.764	0.13
26002/12	16	13	3.625	5.154	-1.529	<.001*
26003/13	13	15	2.615	4.00	-1.385	0.007*
26004/14	13	15	4.154	4.867	-0.713	0.139
27001/11	15	14	2.667	4.571	-1.90	<.001*
27002/12	15	14	3.4	4.071	-0.67	0.142
27003/13	14	15	3.357	3.2	0.16	0.766
27004/14	14	15	2.786	3.667	-0.88	0.079**
27005/15	15	16	3.133	4.125	-0.99	0.077**
27006/16	15	16	3.867	4.938	-1.07	0.048*

Note: two-tailed distribution.
*significant at the .05 level
**significant against one-tailed distribution.

For storyboards 26002/12, 26003/13, 27001/11, 27004/14, 27005/15, and 27006/16, (6 of 10 storyboards) differences in mean ratings to the general appropriateness rating question were observed as predicted. While the other differences in means were not significant, each of these differences in mean, with the exception of storyboard 27003/13, occurs in the predicted direction: the means of breach storyboards are less than the means of their control dual storyboards. The fact that, for 27003/13, the breach storyboard had a higher mean than the control storyboard and that the means for both versions of the storyboard were less than 3.5 suggests that there was some other aspect of the storyboard—besides the breach—that participants were reacting negatively to. The results of the storyboard-by-storyboard hypothesis testing help to shed light on the finding that participants did not recognize the markings aspects of the sequence norm. Participants reacted as predicted to storyboard 27004 and its control dual, however these reactions were tempered by the fact that participants did not react differently to the breach and control versions of storyboard 27003. Reasons why this storyboard did not probe participants as predicted will be considered during the discussion and summary below.

The results of significance testing for Hypothesis II concur with the results of Hypothesis I. These are reported in Table 28, below.

Table 28: Results of significance testing for Hypothesis II: $S_{(I)}A$(breach storyboard)< $S_{(I)}A$(control storyboard).

Storyboard pair (breach/control)	*N* (breach)	*N* (control)	μ_1 (breach)	μ_2 (control)	μ_1-μ_2	*p*
26001/11	16	13	2.313	4.00	-1.688	0.002*
26002/12	16	13	2.313	5.00	-2.688	<.001*
26003/13	13	15	2.615	3.60	-0.985	0.043*
26004/14	13	15	4.00	4.867	-0.867	0.103
27001/11	15	14	2.467	4.643	-2.18	<.001*
27002/12	15	14	3.067	3.857	-0.79	0.167
27003/13	14	15	2.786	3.867	-1.08	0.024*
27004/14	14	15	4.00	4.00	0.00	1.
27005/15	15	16	2.80	4.125	-1.33	0.016*
27006/16	15	16	3.067	5.438	-2.37	<.001*

Note: two-tailed distribution.
*significant at the .05 level
**significant against one-tailed distribution.

In the case of Hypothesis II, for all storyboards in which there is a difference in means (9 of 10), the difference is in the hypothesized direction. Storyboard 27004—which, like storyboard 27003, is about the markings aspect of the sequence norm—showed no difference in means between the *segment of interest* and *distracter segment* rating questions. Possible explanations for why this storyboard, like storyboard 27003, did not elicit the predicted pattern of rating responses will be considered below.

The third across-group hypothesis that was tested for each pair of breach/control storyboards is that there would be no differences in means between the targeted distracter rating questions. Table 29 reports the results of the significance tests for Hypothesis III.

Table 29: Results of significance testing for Hypothesis III: $S_{(D)}A$(breach storyboard) ≠ $S_{(D)}A$(control storyboard).

Storyboard pair (breach/control)	*N* (breach)	*N* (control)	μ_1 (breach)	μ_2 (control)	μ_1-μ_2	*p*
26001/11	16	13	3.563	4.00	-0.438	0.438
26002/12	16	13	4.875	5.077	-0.202	0.576
26003/13	13	15	3.231	4.133	-0.903	0.104
26004/14	13	15	4.462	4.600	-0.138	0.686
27001/11	15	14	4.067	4.643	-0.576	0.179
27002/12	15	14	4.733	5.143	-0.410	0.339
27003/13	14	15	3.857	3.533	0.324	0.517
27004/14	14	15	3.929	4.00	-0.071	0.867
27005/15	15	16	4.733	4.75	-0.017	0.970
27006/16	15	16	5.00	5.00	0.000	1.000

Note: two-tailed distribution.
*significant at the .05 level
**significant against one-tailed distribution.

As predicted, for all pairs of storyboards, there were no significant differences between ratings on the *segment of interest* and *distracter segment* rating questions across breach and control conditions. These rating questions were linked to the common ground across the storyboards in breach and control pairs. The common ground is the 3-5 frame segment of a storyboard that is common to each the breach and control versions of a storyboard. The fact that differences in mean ratings for the *distracter segment* rating questions were not significant contrasts with the results for the *episode appropriateness* and *segment of interest appropriateness* rating questions reported above. For the majority of the storyboards, the mean ratings for the *episode appropriateness* and *segment of interest appropriateness* rating questions for the breach and control segments of the storyboards were lower for the breach storyboards than for the control storyboards. These consistent differences in the predicted direction provide together with the lack of difference for the *distracter segment* rating questions are evidence that participants were reacting to the hypothesized breaches of the semiotic norms, as opposed to other aspects of the storyboards. This point is addressed further during the discussion of the findings.

As reported in Table 29, it is the case that there were no significant differences in means across the breach and control storyboards for the distracter rating questions. However, it is also the case that, for nearly all pairs of storyboards, the means of the breach storyboards were lower than the means of the control storyboards. The overall trend of lower means on the distracter rating questions for the breach storyboards suggests that participants' negative reactions to the moments in the storyboards where a teacher breaches the norm could be affecting their ratings of the common ground part of a storyboard. Such an occurrence would be consistent with the general nature of the reactions that were reported by Garfinkel (1963) during the initial breaching demonstrations. The different kind of reactions to particular moments in the breach and control versions of a storyboard are examined during the analysis of the open response data reported in Chapter 6.

Within-group hypothesis testing

The across-group significance tests reported above indicate that, in general, participants reacted more negatively to storyboards where the teacher is depicted as breaching a norm compared to the control versions of a storyboards (in which there is no breach of a norm). The within-group significance tests aim to provide additional evidence about the differences in reactions to breach compared to control storyboards. Since each participant was asked to rate two specific segments of each storyboard—one where the teacher either does or else does not breach the norm, and one where the teacher does a routine teaching action that is unrelated to the target norm—I hypothesized that, for storyboards where the norm is breached, the means of responses to *segment of interest appropriateness* rating questions (that target the segment of a storyboard in which the teacher breaches a target norm) would be lower than the means of responses to the *distracter segment appropriateness* rating questions (that target the other specific segment of the storyboard participants were asked to rate). This hypothesis is Hypothesis IV. Complementarily, I hypothesized that, for storyboards where the norm is not breached, the means of responses to the *segment of interest appropriateness* rating questions (that target the segment of a storyboard in which the teacher complies with a target norm) would not be significantly different from the means of the responses to the *distracter segment appropriateness* rating questions (that target the other specific segment of the storyboard participants were asked to rate). This Hypothesis V. Hypothesis IV and V are stated for generic breach/control storyboard pairs below.

Within-Group Hypotheses:

IV. H_0: $S_{(I)}A$(breach storyboard) = $S_{(D)}A$(breach storyboard)
H_1: $S_{(I)}A$(breach storyboard) < $S_{(D)}A$(breach storyboard)

V. H_0: $S_{(I)}A$(control storyboard) = $S_{(D)}A$(control storyboard)
H_1: $S_{(I)}A$(control storyboard) $\neq$ $S_{(D)}A$(control storyboard)

Table 30 and Table 31 report the results of significance testing—using paired-samples *t*-tests—for the storyboard-by-storyboard significance testing of hypotheses IV and V.

Table 30: Results of significance testing for Hypothesis IV: $S_{(I)}A$(breach storyboard) < $S_{(D)}A$(breach storyboard).

Storyboard	N	μ_1 ($S_{(I)}A$)	μ_2 ($S_{(D)}A$)	$\mu_1-\mu_2$	p
26001	16	2.313	3.563	-1.250	0.002*
26002	16	2.313	4.875	-2.563	<.001*
26003	13	2.615	3.231	-0.615	0.071*
26004	13	4.00	4.462	-0.462	0.323
27001	15	2.467	4.067	-1.600	0.001*
27002	15	3.067	4.733	-1.667	0.006*
27003	14	2.786	3.857	-1.071	0.046**
27004	14	4.00	3.929	0.071	0.752
27005	15	2.80	4.733	-1.933	<.001*
27006	15	3.067	5.00	-1.933	0.001*

Note: two-tailed distribution.
*significant at the .05 level
**significant against one-tailed distribution.

The statistical tests reported in tables Table 30 show that 8 of 10 breach storyboards prompted significant mean differences between the *segment of interest appropriateness* rating question and the *distracter segment appropriateness* rating question. For the remaining two storyboards, storyboard 26004 shows a non-significant difference in means in the predicted direction, while storyboard 27004 shows a non-significant difference that is opposite of the predicted direction. The fact that Hypothesis IV did not hold for storyboard 27004 provides a further indication that the storyboards that targeted the markings aspect of the sequence norm did not perform according to design.

The results of significance testing for Hypothesis IV are complemented by the results for the significance testing for Hypothesis V, reported in Table 31.

Table 31: Results of significance testing for Hypothesis V: $S_{(I)}A$(control storyboard) < $S_{(D)}A$(control storyboard).

Storyboard	N	μ_1 ($S_{(I)}A$)	μ_2 ($S_{(D)}A$)	μ_1-μ_2	p
26011	13	4.00	4.00	0.000	1.000
26012	13	5.00	5.077	-0.077	0.753
26013	15	3.60	4.133	-0.533	0.150
26014	15	4.867	4.6	0.267	0.217
27011	14	4.643	4.643	0.000	1.000
27012	14	3.857	5.143	-1.286	0.03*
27013	15	3.867	3.533	0.333	0.388
27014	15	4.00	4.00	0.000	1.000
27015	16	4.125	4.75	-0.625	0.155
27016	16	5.438	5.00	0.438	0.203

Note: two-tailed distribution.
*significant at the .05 level
**significant against one-tailed distribution.

Hypothesis IV predicted that mean scores for responses to *segment of interest appropriateness* rating questions would be lower than mean scores for responses to *distracter segment appropriateness* rating questions for storyboards that depict a teacher breaching a norm. Hypothesis V predicted that there would be no such significant differences between rating questions that target specific segments of the control storyboards. The results reported in Table 31 show that, for 9 of 10 control storyboards, this was the case—there were no significant differences between the mean scores of segment-specific rating questions. For storyboard 27012—a storyboard that targets the labels aspect of the sequence norm—the mean ratings on a segment of the storyboard where the teacher complies with the norm is significantly lower than the rating question associated to the common ground segment of this storyboard. Possible reasons for why participants rated the teacher's action in this segment of the storyboard lower than predicted are considered in Chapter 6, where I provide a report of the analysis of the open response data.

Findings of closed-response analysis

The analysis of the responses to the episode and segment specific appropriateness rating questions reported above indicates that participants noticed and reacted negatively to the breaches of the hypothesized semiotic norms. With the exception of the storyboards that targeted the markings aspect of the sequence norm, participants' ratings on the *episode appropriateness* and *segment specific* rating questions for the breach storyboards were lower than participants' ratings on the corresponding questions for the control versions of the storyboards. These findings are consistent with the results of the norm-by-norm significance testing that was reported in the first part of Chapter 5. Furthermore, the responses to the *distracter segment appropriateness* rating questions showed less variation across the conditions.

Altogether, the results reported above provide evidence that there are normative ways that semiotic resources are used when doing proofs in geometry. The recognition of the *less* and *more* details norms gives an indication that the degree to which teachers expect that students will "show every step" depends on whether or not the step that is shown is something that could be assumed from the given diagram, versus being entailed by a written statement. The recognition of the labels and reasons aspects of the sequence norm gives an indication that there is nothing out of the ordinary about students creating mark-for-mark reproduction of already completed proofs when they are called to the board to present their work. The transcription of the different resources of the proof can proceed in the order that makes the transcription work convenient.

The overall trends in the data reported above suggest that participants noticed breaches of semiotic norms and, in turn, that the virtual breaching experiment with control performed according to how it was designed. The responses to storyboards that targeted the markings aspect of the sequence norm are set apart from the overall pattern. For storyboard 27003, there were differences as predicted between the *segment of interest appropriateness* ratings, but there was no difference between participants' *episode appropriateness* ratings on the storyboards. For storyboards 27004 and 27014, the *episode appropriateness* ratings of the breach storyboard was significantly lower than the *episode appropriateness* rating for the control storyboard, but there were no differences between the *segment of interest appropriateness* ratings. The lack of recognition of breaches to the markings aspects of the sequence norm suggests that markings have a different status than labels and reasons during the work of presenting a proof.

One possible explanation for these unexpected results could have to do with the role that markings play in the written argument of a proof. Unlike labels—which anchor the written statements in a proof to its accompanying diagram—or reasons—which provide the justifications that warrant each step in a sequence—markings are auxiliary visual features that are applied to a diagram that accompanies a proof. Markings could thus be seen as having less value than labels or reasons. If this were true, then the mixed results of the analysis of responses to rating questions for the breach and control versions of the markings storyboards would be an indication of the relative weakness of the underlying signal. That is: Markings and the role that markings play in a proof are less noticeable than labels—without which the statements of a proof are senseless—or reasons—without which the statements in a proof are unjustified—and therefore participant reactions to a breach of the markings aspect of the sequence norm were more diverse. This is considered more carefully in Chapter 6.

The next section of the chapter reports the results for the analysis for the two sentence completion questions. I provide contingency tables for this data, aggregated by norm and sorted by choice and report the results of statistical testing using binomial tests. The last section of the chapter considers what the responses to the sentence-completion tasks indicate about the virtual breaching experiment with control instrument.

Analysis of the justification data

In addition to the rating responses analysed above, participants answered two closed-ended sentence completion questions. These questions asked participants: (1) "To justify the way the teacher reviewed the work on the board, I would say the teacher…" (justify) and (2) "To criticize the way the teacher reviewed the work on the board, I would say the teacher…" (criticize). Participants were given 5 choices following each sentence and forced to choose one—and only one—of the five choices. The choices were the same for each type of question: "To justify the way the teacher reviewed the work on the board, I would say the teacher…"

- appropriately represented the mathematical concepts and properties under consideration (Disciplinary)
- appropriately attended to the individual needs of students (Individual)
- appropriately supported the social interaction of the class (Interpersonal)
- appropriately complied with the policies and practices common in schools (Institutional)
- I would not justify what the teacher did (None)

For the *criticize* version of the question, the words "did not" were inserted before appropriately for the first four choices and *justify* was changed to *criticize* for the last choice.

The purpose of including the sentence completion questions was to provide participants with another avenue for reporting their reactions to the storyboards. The justify/criticize questions used in this study are adapted from previous instruments in which teachers viewed and rated representations of classroom situations[47]. The four justifications provided are conceptually grounded in what Herbst and Chazan (2011) have described as the *professional obligations* of mathematics teaching.

Herbst and Chazan (2011) identify the discipline of mathematics (disciplinary), the individual student (individual), the social interaction of a group students (interpersonal), and the institution of schooling (institutional)—including local, regional, and national policies and procedures that constrain the work teachers do in classrooms—as stakeholders in what takes place in mathematics classrooms. They argue that professional mathematics teachers are obligated to these stakeholders: To the extent that mathematics teachers are professionals, what teachers do in classrooms is accountable to the stakeholders identified above.

For example, were a teacher to decide to depart from the way that a mathematical concept is presented in the curriculum, and were the teacher asked to give an account of this departure to an agent with vested interest in the school—such as a colleague, supervisor, or parent—it would be incompatible with the notion of the teacher as professional for the teacher to admit that "I just felt like doing something different" or "the definition in the textbook confused me". Regardless of whether such idiosyncratic motives influenced what the teacher actually did, part of assuming the identity of "mathematics teacher" is an awareness that such accounts are unprofessional. Instead, what a professional mathematics teacher in such a situation is likely to offer is an account that invokes one of the four professional obligations identified above—e.g. "I used a more general definition than what is provided in the book, because I wanted the students to see how the ideas cohere from an algebraic standpoint" (disciplinary obligation), or: "Students

[47] A study by Herbst, Aaron, Dimmel, & Erickson (2013) used the expressions "favorably rate" and "unfavorably rate" in the questions stems and used *not applicable* as the fifth choice.

generally get confused by the book's presentation. The way that I formulated the idea is easier for students to understand" (individual obligation).

Herbst and Chazan (2012) posit that the professional obligations could be a source of warrants teachers might invoke to justify departures from normative classroom practice. In the case of the virtual breaching experiment with control instrument, the sentence completion questions were provided to give participants generic statements of such warrants. While more specific information about how participants might justify or criticize the teaching they witnessed in the different storyboards comes from responses to the open-ended questions—the analysis of which is reported in Chapter 6—the sentence completion questions provide coarse measures of the grounds on which the appropriateness of the teacher's actions were assessed.

I scored the responses to the sentence-completion questions by coding (1) for the rationale selected by a participant and (0) for the outcomes that were not selected. I totaled the choices for each of the 5 rationales by storyboard. Table 32 aggregates the results of this scoring across all storyboards in the breach and control conditions. I report aggregate responses to provide an instrument-wide view of how participants reacted to the breach and control versions of the storyboards. This wider view of the storyboards complements the norm and storyboard specific results that were reported above.

Table 32: Total number of justifications selected for the *justify* or *criticize* questions, by question (*justify,* criticize) and condition (breach/control)

Question/ condition	Disciplinary	Individual	Interpersonal	Institutional	None	*N*
Justify/Breach	39	10	17	13	67	146
Justify/Control	48	27	26	17	28	146
Criticize/Breach	31	61	24	6	24	146
Criticize/Control	13	33	29	1	70	146

I modelled the column-totals for each table as the number of successful outcomes of 5 mutually exclusive binomial events. Let k_i be the sum of cell-counts in column *i.* Then *i* ranges from {1-5}, where {1 = Disciplinary, 2 = Individual, …, 5 = None}. If the choices were distributed randomly, the expected value, e_i, for each k_i is 29.2[48]. I performed binomial tests[49] against the null hypothesis that the outcomes would be randomly distributed among the 5 choices (i.e., for each the five binomial events, the expected number of successes is 29.2). The results of this significance testing are reported in table Table 33.

Table 33: Binomial tests for contingency table totals, by condition and question

Binomial Tests for Contingency Table Totals, by Condition, Question

	Breach				Control			
	Criticize	*p*	Justify	*p*	Criticize	*p*	Justify	*p*
Disciplinary	31	0.311	39	0.019**	13	0.000**	48	0.000**
Individual	61	0.000**	10	0.000**	33	0.186	27	0.369
Interpersonal	24	0.166	17	0.005**	29	0.467	26	0.293
Institutional	6	0.000**	13	0.000**	1	0.000**	17	0.005**
None	24	0.166	67	0.000**	70	0.000**	28	0.451
Total:	146		146		146		146	

**Significant at .05 level

The results reported in Table 33 indicate the numbers of warrants that were provided in each category (e.g., disciplinary, individual…) in response to the *justify* and *criticize* sentence completion question. The two columns on the left of the table report the aggregate results of the sentence completion questions for the breach storyboards, by question type. The two columns on the right of the table report the aggregate results of the sentence completion questions for the control storyboards, by question type. Because results that were significantly greater than or significantly less than the expected value would be significant when tested against the binomial distribution, the reported *p*-values are standardized according to the following procedure: If $k_i < e_i$, then the reported *p* value is the cumulative binomial probability. If $k_i > e_i$, then the reported *p*

[48] This is 146/5.
[49] Tests performed in Microsoft Excel.

value is 1-{cumulative binomial probability}. This procedure is equivalent to reporting the cumulative probability for *at most k* successes in *n* trials, whenever $k_i < e_i$, and *at least k* successes in *n* trials, whenever $k_i > e_i$. Reporting standardized *p* values helps to group results that are significant at the .05 level.

The main result reported in Table 33 is the complementarity between the answers to the *criticize* and *justify* questions in the breach and control conditions. In the breach condition, the most frequent justification is *none*, meaning that participants were most likely to complete the justify sentence for these storyboards with: “I would not justify what the teacher did.” On the other hand, the most frequent criticism of the control storyboards is also *none*, meaning that participants were most likely to complete the criticize sentence for these storyboards with: “I would not criticize what the teacher did”. Each of these response rates (i.e., the *none* option for the breach-justify question and the *none* option for the control-criticize question) differs significantly from what one would expect if the outcomes were randomly distributed among the 5 choices, as reported in Table 33 (p <.001 in each case). Similarly, the response rates for the *none* option on the breach-criticize (meaning: “I would not criticize what the teacher did”) and control-justify (meaning: “I would not justify what the teacher did”) sentence completion tasks do not differ significantly from what would be expected by chance.

The data reported in Table 33 provides additional evidence that the virtual breaching experiment with control instrument performed according to design. In particular, the differences in warrants to justify or criticize the work of the teacher in breach and control storyboards indicate the range of ways that participants reacted to departures from the expected classroom routines. One area where this is evident is in the use of the *individual* warrant as a justification or criticism.

Each breach-control storyboard pair featured students that could be seen to be put in a difficult situation by the teacher—i.e., one could find students or student work that are scrutinized or dismissed by the teacher in the breach and control versions of each storyboard. Yet looking at the counts for the warrants, the individual warrant was chosen to criticize the teacher 61 times in the breach storyboards, compared to only 33 times in the control storyboards. Looking at this warrant from the other direction, the individual warrant was selected 10 times to justify the work the teacher did in the breach storyboards, compared to 27 times for the control storyboards. The totals for the breach storyboards—61 (criticize) and 10 (justify)—would occur by chance alone at rates that are much lower than .05: In each case, $p < .001$. While for the control storyboards, the total outcomes are within the range of what one would expect if the outcomes were distributed randomly: $p = .163$ and $p = .401$, respectively.

The responses to the sentence-completion questions provide evidence that the breach and control storyboards are recognizable as episodes of instruction. The second-most frequently occurring response to the breach-justify sentence completion question is the *disciplinary* warrant ($n = 39$, $p = .019$), meaning participants indicated that: “the teacher appropriately represented the mathematical concepts and properties under consideration” at a rate that was significantly different from what would be expected by chance alone. The fact that the *disciplinary* warrant was the second-most frequently selected choice for grounds on which the teacher’s work could be justified in the breach storyboards suggests that participants were able to recognize the mathematical reasonableness of the teacher’s work in spite of the breach. Conversely, the use of the disciplinary warrant to criticize the work of the teacher in the breach storyboards did not differ significantly from what would be expected from chance ($n = 31$, $p = .311$). In the case of the control storyboards, the disciplinary warrant was the most frequently-selected warrant for justifying the work that the teacher did (n= 48), and this warrant was selected at a rate that differs significantly from what would be expected by chance (p <.001).

The differences reported above in the warrants that participants selected to justify or criticize the work of the teacher suggest that the segments of the storyboard that depict a teacher breaching a hypothesized semiotic norm are affecting what participants' are seeing in the depicted classroom situations. The differences across the breach and control versions of a storyboard are minor from norm to norm: In each version of the storyboard, the same general things happen and the teacher does many of the same things. Yet the 3-5 slide segment where the teacher departs from the hypothesized norm in each storyboard set are sufficient to produce the differences in responses to the rating and sentence completion questions reported above. The analysis of the general warrants reported above is a first approximation of reactions to the storyboards that are more specific than the appropriateness ratings of the work of the teacher at different segments in the storyboards. The continuation of these more specific analyses is the focus of Chapter 6, which reports on the analysis of the open response data.

Chapter Summary

In this chapter, I reported the results of the quantitative analysis of the two kinds of closed-response data. The first part of the chapter presented the analyses of the responses to the closed-ended rating questions. The planned comparisons across and within experimental groups were then conducted using independent and paired t-tests. The results of the inferential analysis provide evidence that the hypothesized effects—namely: lower scores on the general and targeted appropriateness rating questions for the breach compared to the control versions of a storyboard—were observed as predicted. The final section of the chapter reported results of the analysis of the sentence-completion questions. The differences in selected warrants across the breach and control versions of the storyboards provided an additional source of evidence that participants noticed and reacted to breaches of hypothesized semiotic norms.

CHAPTER 6:

WHAT DO TEACHERS REPORT ABOUT BREACHES? AN ANALYSIS OF OPEN RESPONSE DATA

I describe here the analysis of participants' responses to open-ended questions on the virtual breaching experiment with control. The open response data provides a perspective on participants' reactions to the breach and control versions of the storyboards they viewed that complements their responses to the closed-ended questions. The first part of the chapter presents examples of open responses to the various storyboards. These example responses serve to illustrate the range of responses to the different storyboards. Following the presentation and discussion of these examples, I describe the schemes that I used to code the open response data for evidence that participants recognized the breach (or non-breach) of the semiotic norms at stake in the storyboards or made evaluative comments about the storyboards. The final part of the chapter reports on the mechanics of the open-response coding, the reliability study of the coding schemes, and the results of the analysis of the coded open-response data.

Overview of the open-response data

After viewing each storyboard, participants were given four opportunities to provide open-response data. The first question that participants were asked is: "What did you see happening in this scenario?". The purpose of prompting participants with this broad, open-ended question was to capture participants' overall reactions to the instances of doing proofs (hereafter: situation instances) that were represented by the different storyboards. In classic breaching experiments, the experimenter creates the breached situation by enacting a deliberate departure from the expected social order. As a direct participant in creating this breached situation, the experimenter can observe directly how the other participants in the situation are responding to the breach. Since participants in the study reported here were viewing representations of teaching situations—rather than participating in those situations directly—the general open-response question provides a virtual analog of witnessing or observing how a person responds to the situation. This general open response question has been used in previous virtual breaching experiments (Herbst, Aaron, Dimmel, & Erickson, 2013). Examples of responses to this question for each pair of breach/control storyboards are reported and discussed in the next section.

Participants had three other opportunities to provide open responses, following their review of each storyboard. These open response fields followed the episode—how appropriate was the teacher's review of the proof in this scenario?—and segment-specific—how appropriate were the teacher's actions in this segment of the storyboard?—appropriateness rating questions (the analysis of which was reported in Chapter 5). Following each rating question, participants were prompted to "Please explain your rating." Participants were required to enter text into this explanation field—in order to continue to the next screen in the online experience that hosted the instrument—though "no comment" was entered for 16 of the 342 follow-up responses (4.67% or responses). The purpose of these follow-up open response fields was to provide participants with opportunities to make additional comments about the salient features of the situation instances that were represented in the storyboards.

The four open response fields presented participants with opportunities to comment on the representations of situation instances with varying degrees of focus. The first open response prompt—What did you see happening in this scenario—probes for participants' reactions to the storyboards at the most general level. The next question is the open-response prompt that follows the general rating question—how appropriate was the teacher's review of the proof? This question attempts to focus participants' ratings on the actions the teacher completed to review the proof, however the "review of the proof" includes other moments in the storyboard besides those where the teacher breaches—or fails to breach—the target norm. For example, how a teacher responds to a student who proposes an alternate method for completing the proof, or a how a teacher answers a question about how to differentiate the substitution from the subtraction properties of equality are actions taken by teachers that are unrelated to the target semiotic norms. When participants respond to these general rating questions, they could be keying in on how the teacher in a storyboard conducts these other tasks of teaching. The open response field that follows the first rating question thus provides an opportunity for the participant to link their rating to a specific segment of the storyboard. Since this open response field follows a rating question—and could therefore be seen as priming the participants to provide an evaluative comment—it is not at the same level of generality as the first open response question, though is still more general than responses that follow the specific rating questions.

Each of the segment-specific rating questions asks participants to rate the appropriateness of the teacher's actions during a specific segment of the storyboard. The *segment of interest* rating question shows the participants a 3-to-5-frame segment of the storyboard, during which the teacher either breaches or does not breach the target norm. The *distracter segment* rating question shows the participants a 3-5 frame segment of the storyboard during which the teacher does some other act of teaching. These rating questions focus participants' responses on specific moments of the storyboard and the responses that participants provided in the follow-up fields are therefore the least general of the four open-response fields the participants provided. In the next section, I provide examples of responses to each of the different fields for breach and control storyboards. Following this presentation and discussion of these examples, I describe the methods that I used to code the open responses for evidence that participants repaired the situations that were represented by each storyboard.

Examples of open responses

Participants provided a range of responses to the various open-response fields. The first part of the chapter presents examples of open responses that indicate this range for the various storyboards. The goal of beginning with example responses is to provide some motivation for the coding schemes that were developed and applied to the open response data. The coding schemes and the protocol by which they were applied to the data are described later in the chapter.

Responses to the first open response question

The responses to the first open response field are the closest approximation to simply witnessing how participants negotiate, or react to, a social situation. The responses to the first open-ended question have the potential to carry the strongest signal that participants noticed—or failed to notice—when a norm of the situation had been breached. It is reasonable to expect that responses to the first question following storyboards that breached a target norm were more likely to be negative than responses to the first question following storyboards that did not breach the norm (see chapter 4). The responses included below were selected to present examples and non-examples of responses—by breach/control storyboard pair—to the first open response field that exhibit this response pattern.

Responses to the *less details* breach and control storyboards.

There were two storyboards that breached the details norm by showing a teacher accept a proof that I hypothesized would be viewed by teachers as having insufficient details in the written argument of the proof. In storyboard 26001, the teacher accepts a proof in which the congruence of two segments is not explicitly deduced from the definition of midpoint. In storyboard 26002, the teacher accepts a proof in which the distinction between *angles* and *measures of angles* is not strictly observed in the statements and reasons that accompany the proof. In the control duals of these storyboards, the teacher behaves in what is hypothesized to be the normative way: Rather than accept these proofs, the teacher insists that the proofs are not complete until these additional details are stated in the proof.

The responses that participants provided to the first open response field indicate that participants noticed and reacted to the breach of the norm in each storyboard. For example, in response to the "what did you see happening…" question following storyboard 26001, a participant wrote: "*The teacher did not correct the student for skipping the statement in the proof* - proving sides congruent." (participant 2317, emphasis added) The participant characterizes the action that the teacher took in the scenario as *not correcting the student.* The participant describes *skipping the statement in the proof* as an error that the teacher should have corrected. This response indicates that the breach of the details norm—by the teacher cashing a proof that does not have required details explicitly stated in the statements and reasons of the proof—was noticed by the participant.

A second response to storyboard 26001 indicates not only that the participant noticed the breach of the norm, but also offers an explanation of why this action by the teacher is questionable: "…I don't like that *the teacher said you didn't have to include the segments being congruent in the proof.* You would have to include that to prove two triangles are congruent. A proof shouldn't leave *any piece up to the imagination or interpretation*." (participant 2020, emphasis added). As in the previous example, this response provides evidence that the participant noticed the breach of the less details norm. The participant also states that, "A proof shouldn't leave any piece up to the imagination or interpretation." This response provides evidence that teachers recognize some details as being necessary to state explicitly even if they are entailed conceptually by other statements—in this instance: the congruence of the two segments is entailed by the definition of midpoint. In the case of the more details storyboards—described below—there are examples of responses that make an almost exactly opposite argument: The teacher in these storyboards is criticized for being too focused on the minutia at the expense of the overall view of the proof.

Responses to the "what did you see happening in this scenario" question for storyboard 26002 also provide evidence that participants noticed the breach of the details norm by the teacher cashing a proof that omitted some expected details. Participant 2411 provides an example response: "The teacher *was not concerned about details about 'm' for measure* and did not check for understanding after the subtraction/substitution discussion." (emphasis added) This response indicates that the teacher was "not concerned" about the details of writing *m* before the "<" symbol to indicate that the measures of the angles (as opposed to the angles themselves) are what is being added. Furthermore, this participant also raises an issue with the teacher's handling of the distinction between substitution and subtraction, perhaps an indication that the participant's negative reaction to the breach is influencing perception of other parts of the storyboard. A different participant writes: "The teacher is *down playing the little things*. Sometimes those *little things can change the whole outcome.*" (participant 2300, emphasis added). This response indicates that the teacher's "downplaying" of "the little things"—which seems most likely to be a reference to the fact that the teacher does not insist on explicitly recognizing the distinction between angles and their measures in the written record of the proof—is questionable because those little things could "change the whole outcome". Like the responses to storyboard 26001, these responses indicate that participants recognized the breach of the details norm in the storyboards where the teacher is depicted as accepting a proof that is missing some of the expected details.

Responses to the control versions of the less details storyboards provide evidence that the storyboards represented the routine ways that teachers insist on certain kinds of details when checking proofs in geometry. For example, one response to storyboard 26011 reads: “The teacher asked a student *(who had an incomplete proof)* to show his work on the board. I assume that the teacher picked that student *to learn from their incorrect answer.*” (participant 2203, emphasis added). This response provides evidence that the participant recognized that the student proof—that was missing a hypothesized to be expected detail—was not correct. The participant describes the fact that the proof is not correct as a learning opportunity for the other students in the class. A different participant responded to the same storyboard by writing: “I saw amazing geometry students who were actively participating in the class problem.” (participant 2313) While this response does not provide any specific reference to the missing steps in the proof, the overall positive assessment of the situation in this response suggests that the teacher’s insistence on providing the additional details of the proof were not unexpected.

Responses to the control version of storyboard 26002 also provide evidence that insisting on certain kinds of details is routine when checking proofs in geometry. Participant 2382 wrote: “The teacher did a great job of giving them *a problem that needed to be modified.*” (emphasis added) In this example, “providing the students with a problem” refers to the fact that the teacher presents the students with a proof that “someone from 5th hour” did. According to the participant’s response, that this proof had a perceived mistake in terminology—failing to maintain the distinction between angles and their measures—is the reason that it needed to be modified. And because the teacher provided the students with a problem that needed to be modified, the teacher is recognized as having done a "great job." Similarly, participant 2333 wrote: "Good interaction between students and teacher. Good communication and discussion of *proper steps of a proof.*" (emphasis added) This participant recognizes that the teacher's insistence that the steps in the proof be revised to maintain the distinction between *angles* and *measures of angles* is tantamount to the teacher exhibiting the proper steps of the proof.

These example responses to the breach and control versions of the *less details* storyboards provide evidence that participants noticed when the teacher breached the details norm by accepting a proof that omits some of the expected details. Furthermore, there is also evidence in these responses that participants objected to the teacher's actions on the grounds that valid proofs "make no assumptions" or "take nothing for granted." The responses to the *more details* storyboards approach the issue of the necessary level of detail from the opposite direction. In these storyboards, the teacher breaches the *details norm* by insisting that students provide details that typically are not required in geometry classrooms, though as is the case with the *less details* storyboards, the information the teacher insists on is mathematically warranted. Responses to the *more* details storyboards are presented below.

Responses to the *more details* breach and control storyboards

There were two storyboards that breached the details norm by showing a teacher insist on including details in a proof that are not usually required when doing proofs in high school geometry. In storyboard 26003, the teacher insists that a student unpack a conclusion that two angles are supplementary into more fundamental steps. In the breach version of the storyboard, the student is required to deduce that the angles are supplementary by first establishing the collinearity of the points that define the base rays of two adjacent angles. In the control version of the storyboard, the student is still required to unpack the statement into more primitive steps, but the collinearity of the points is taken for granted. The breach version of storyboard 26004 shows a teacher problematizing the existence of a point of intersection of the angle bisectors of two consecutive angles of a parallelogram. In the control version of this storyboard, the teacher proceeds with the proof while taking the point of intersection of the angle bisectors as given by the diagram.

In each of these storyboards, the teacher's insistence on these additional details is consistent with the idea that, in a proof, there can be nothing that is taken for granted—a view that is evident in the example responses cited above. However in the open responses to the "what did you see happening in this scenario?" question for each of the more details storyboards, there is evidence that participants reacted negatively to the teacher for, essentially, making much ado about nothing. For example, in one response to storyboard 26003, participant 2203 writes: "A teacher making a example of student for "skipping steps." This participant goes on to say, in the same response, that: "(*I feel those steps aren't as important,* and rather than have the student build confidence in his proof-writing ability, the teacher has made that one student feel as if he can't write a proof.)" A different participant writes: "A 99 step proof? *This teacher is being a bit ridiculous on the thoroughness of the proof.* The proof is getting lost in the minutia. If the teacher is that concerned about the collinearity of points, then include it in the given statement." (participant 2333, emphasis added). The specificity of these answers to the most general open-response question indicate the salience of the teacher's insistence on—what are characterized by participants as—"less important" details for participants.

The responses to the "what did you see happening in this scenario?" question to storyboard 26004 also indicate that participants reacted negatively to the teacher's insistence that the existence of the point of intersection of the angle bisectors be explicitly stated in the proof. Participant 2203 wrote: “The teacher was trying to sound smarter than the book he was using; making his students feel as if they "don't get it." Rather than solving the problem, *he was trying to pick apart the details of the given picture.*” (emphasis added) A different participant offers a less specific, though still negative, assessment of the teacher’s actions in the scenario: “I saw a teacher who was unprepared. She wa[s]ted time writing the problem on the board, (it should've already been up there), and *she did not properly teach the proof*” (participant 2248, emphasis added). This response is not as specific as the other example responses, yet it still indicates that the participant reacted negatively to how the teacher presents the proof in the scenario. The statement that “the teacher did not properly teach the proof” suggests that there was some aspect of the teacher’s presentation that the participant found objectionable. While this participant does not offer specific language about the teacher’s insistence on establishing the existence of the point of intersection, this response nonetheless suggests that an approach to teaching the proof that includes such action by the teacher is “not proper”.

As was the case for the *less details* storyboards, the control versions of the *more details* storyboards provide evidence that the storyboards represented the routine ways that teachers scrutinize steps when checking proofs in geometry. After viewing storyboard 26013, participant 2236 wrote: “The teacher helps the student to technically solidify her proof by asking her additional questions.” This comment describes the teacher’s work in the scenario as helping a student to technically solidify her proof. Of this same storyboard, participant 2001 wrote: “A teacher helping to guide students through a proof.” This response suggests that the teacher’s participation in the student presentation of the proof—during which the teacher asks a student to unpack the fact that two angles are supplementary into two distinct steps, one that indicates the angles are a linear pair, and one that concludes they are supplementary, by the supplements theorem—was not unexpected (and therefore, unremarkable). That the teacher handled the presentation of the proof in routine ways is also evident in the responses that participants provided to the "what did you see happening in this scenario?" after they viewed storyboard 26014. For example, participant 1514 responded by writing: "The instructor is helping the students think through the proof." Of this same storyboard, participant 651 wrote: "teacher leading a class thru a proof." Each of these responses suggests that the teacher behaved as expected when the teacher did not explicitly establish the existence of the point of intersection of the angle bisectors.

Other responses to the *details* norm storyboards

Not all responses to the first open-ended question for the storyboards that breach the details norm provide evidence that participants noticed or reacted strongly to the breach of the norm. For example, in response to storyboard 26001, one participant wrote: "Discussion of a proof and what someone did compared to what they did. Use of different methods" (participant 2303). Of this same storyboard, a different participant wrote: "The teacher is explaining two different ways to do the same problem." (participant 2311) Similarly, in response to storyboard 26002, one participant wrote "Teacher shared with current class proof that was used in previous class and asked students to make comments on it." (participant 2056) Of this same storyboard, a different participant wrote: "Whole class discussion centered around a geometric proof." (participant 2408). Comparable examples of unremarkable assessments of the teacher's work are evident in the responses to the breach versions of the *more details* storyboards. How these variations in open responses were analyzed is reported below.

Responses to the *sequence (labels)* breach and control storyboards

There were two storyboards that depicted breaches of the labels aspects of the hypothesized sequence norm. In storyboard 27001, a student transcriber forgets to label the points on the diagram, even though these labels are used in the statements and reasons of the proof that the student transcribed. Rather than let the student add the labels after the proof (with statements and reasons) has been written on the board, the teacher labels the diagram for the student, using labels that do not match what is in the statements and reasons that are already written. According to what the teacher says to the student in the storyboard, the teacher does this to ensure that the student is thinking about the connection between the labeled diagram and the statements and reasons as the proof is being written on the board. I hypothesized that the teacher's actions in this storyboard would be viewed by participants as breaching the sequence norm, because the teacher does not permit the student to simply add in the labels after the transcription has been completed. Allowing the student to add the labels would comply with the hypothesized norm and is the action that occurs in the control dual of storyboard 27001.

In storyboard 27002, a student transcriber is in the process of reproducing a proof from a sheet of paper onto the classroom whiteboard. The student has included several statements that refer to numbered angles, but the student has not yet added those labels in the accompanying diagram. A different student in the class asks what the numbers in the statements refer to. In the breach version of the storyboard, the teacher breaches the sequence norm by asking the student at the board to stop the transcription, erase what has been done already, and start over, this time labeling the diagram before using those labels in the statements of the proof. I hypothesized that the teacher's actions would be viewed by participants as breaching the sequence norm, because the teacher asks the student to modify the proof before the student has completed the transcription. In the control dual of storyboard 27002, the teacher complies with the hypothesized norm by telling the student that asks about the labels to hold the question until the student at the board has completed writing up the proof.

Participants responses to the "What did you see happening in this scenario?" question following the breach versions of storyboards 27001 and 27002 cast the teacher's interference with the student transcription of the proof as unacceptable. Following storyboard 27001, a participant wrote: "I saw a teacher force a student to go through unnecessary steps to satisfy some unknown rationale." A different participant wrote more generally about the teacher's actions in that same storyboard: "I saw the teacher being mean to a student." These example responses to the breach version of storyboard 27001 illustrate the general negative way in which participants responded to the teacher's actions in this storyboard.

In addition to these generally negative responses, there were responses that more directly recognized the sequence norm. One participant wrote: "I think the teacher created unnecessary work for the student. They labeled their paper with the correct labels and overlooked them on the board." This response indicates that the work that the student was doing on the board had been completed already on the student's paper. A different participant wrote: "pointing out that she needed to number the angles was enough. No need to spend the time rewriting the whole thing." This response indicates that it would have been sufficient to point out to the student that the angles needed to be numbered, suggesting that it would have been adequate for the student to add in the labels after the proof had been written on the board.

Responses to the control versions of storyboards 27001 and 27002 suggest that the student's presentation of the proof to the class was routine. One participant wrote: "This played out just like I'd want it to, I think. Teacher asking the student to put something on the diagram they forgot, then asking questions and helping with the flow of the conversation without dictating it...” (27001BC131-Control). This response indicates that it was not necessary for the diagram to be labeled before the student used the labels in the written statements and reasons of the proof. The labels were simply a part of the record of the proof that the student “forgot” when writing the proof on the board. Other responses looked past the student presentation of the proof altogether. For example, one participant wrote: “the teacher is validating how different methods are acceptable” (27001BC181-Control). This is an example of a response which suggests that, for the participant who wrote this response, the student’s presentation of the proof was part of the background expectancies of the classroom action that took place. Exemplifying the routine nature of the student’s presentation, a different participant wrote: “nothing out of the ordinary” (27001BC11-Control).

Responses to the *sequence (markings)* breach and control storyboards

There were two storyboards that depicted breaches of the the markings aspect of the hypothesized sequence norm. In storyboard 27003, a student, Gamma, volunteers to put a proof on the board. As the student copies the proof onto the board, the student marks angles in the diagram as congruent before writing the statements and reasons that establish that the angles are congruent. Two students in the class ask notice this and ask the teacher about the reasons for the angle congruence that Gamma marked on the diagram. The teacher asks Gamma to erase the markings in the diagram and instructs Gamma to add them back once the statements and reasons that establish the congruence of the angles have been written in the proof. I hypothesized that the teacher’s actions in this storyboard would be viewed by participants as breaching the sequence norm, because the teacher is insisting that the markings in the diagram be coordinated with the statements and reasons of the proof. In the control dual of this storyboard, the teacher tells the students that asked about the markings to hold their questions until Gammas has finished writing the proof on the board.

In storyboard 27004, a student, Lambda, has just finished writing a proof on the whiteboard from a homework worksheet. Lambda marked the diagram on the worksheet but did not mark the diagram on the whiteboard. The teacher notices this and asks Lambda to come back up to the board and verbally explain the proof, adding markings to the diagram as the those properties become established by the statements and reasons of the proof. I hypothesized that the teacher's actions in this storyboard would be viewed by participants as breaching the sequence norm, because the teacher does not recognize that Lambda was simply copying the proof that was already written on the worksheet onto the whiteboard. In the control dual of this storyboard, the teacher complies with the hypothesized sequence norm by permitting Lambda to simply add the markings to the diagram.

As was the case for the storyboard that target the *labels* aspect of the *sequence* norm, participants' responses to the "What did you see happening in this scenario?" question following the breach versions of the storyboards that target the *markings* aspect of the *sequence* norm take issue with how the teacher interacts with the student presenting the proof to the class. One participant wrote: "Teachers makes kid at the board rework their problem. (waste of time)" (27003BC221-Breach). This response indicates that the teacher "makes" the student "rework" the problem and states that it is a "waste of time" for the student to be required to do so. A different participant wrote: "A student was putting up a homework problem on the board which she copied from here (*sic*) paper and the teacher overreacted to her not putting marking (*sic*) on the figure and insulted her." (27004BC161-Breach). Here, the participant acknowledges that the student is copying a homework problem from her paper onto the board. The issue for the participant is not with the student who is doing the copying, but rather with the teacher for overreacting and insulting the student.

Responses to the control versions of the storyboards that target the *markings* aspect of the sequence norm provide evidence that the student's transcription of the proof problem onto the board was a routine way for a student to present a proof. For example, one participant wrote: "The teacher allowed the students to work together and discuss key topics" (27003BC231-Control). A different participant wrote: "I saw a student with incomplete work and a teacher who encouraged the student to show all work completely." (27004BC51-Control). Each of these responses indicates that the student's presentation of the proof unfolded in such a way that it was not worth remarking on when answering the general open response question.

Responses to the *sequence (reasons)* breach and control storyboards

There were two storyboards that depicted breaches of the the reasons aspect of the hypothesized sequence norm. In storyboard 27005, a student is transcribing a proof onto the board. The student has finished, and the teacher notices that the student did not write a reason for one of the steps in the proof. The teacher points this out to the student. The student realizes which step is missing a reason. The student moves to add this reason to the poof, but the teacher insists that the student erase all of the steps below the step that was missing the reason, and to redo the proof from that point, linking the statements and reasons as the proof is being written on the board. I hypothesized that the teacher's actions in this storyboard would be viewed by participants as breaching the sequence norm, because the teacher does not allow the student to add the missing reason to the transcription of the proof. Instead, the teacher is insisting that the student re-produce the proof on the board in a logically coherent way. In the control dual of this storyboard, the teacher permits the student to add the missing reason to what is already written on the board.

In storyboard 27006, a student, Omicron, volunteers to put a proof on the board. As the student is transcribing the proof, the student writes all of the statements before writing any of the reasons that support those statements. Two different students in class notice this and ask the teacher how Omicron knows those statements are true. The teacher picks up on the student questions and asks Omicron to re-do the proof, this time, writing the statements and reasons in sequence. I hypothesized that the teacher's actions in this storyboard would be viewed by participants as breaching the sequence norm, because the teacher insists that Omicron reproduce the proof on the board in a logically coherent order. In the control dual of the storyboard, the teacher tells the students that have the question about Omicron's work to wait until Omicron has finished transcribing the proof.

The storyboards that target the *reasons* aspect of the *sequence* norm also suggest that different segments of the storyboard were salient for participants depending on whether participants viewed the breach or control version of a storyboard. The breach versions of the storyboard prompted responses that suggested the teacher's handling of the student's presentation of the proof was unexpected. One participant wrote: "Teacher needlessly wasted time, by making a correct student erase and rewrite his proof in the order that the teacher preferred." (27006BC211-Breach). Along these same lines, a different participant wrote: "A student who put themselves forward as willing to display their work in front of their peers and although doing a good overall proof, the teacher did not support nor encourage the student in their efforts" (27005BC291-Breach). Both of these responses indicate that the student's presentation of the proof is acceptable and that the teacher was, at best, failing to encourage the student and, at worst, conducting a needless waste of time.

Responses to the control versions of the storyboards that target the *reasons* aspect of the *sequence* norm suggest that the student's presentation of the proof and the teacher's interaction with the student during the presentation is routine. One response that illustrates the perceived-to-be routine nature of the teacher's actions is: "The teacher is checking the student's work. Other students are comparing their work to hers" (27005BC171-Control). Another participant wrote: "students were discovering that there are different ways to prove things" (27006BC261-Control). In each of these example responses, the manner in which the student presents the proof to the class is unremarkable.

Other responses to the *sequence norm* storyboards.

There were responses to the breach and control versions of the sequence norm storyboards that do not fit the profile of the example responses considered above. Some participants took issue with the fact that the teacher did not require the student to explain the proof as it was being written on the board: "why didn't the teacher have the student explain each step and talk about why it was correct or not" (27005BC261-Control). This is a response to the control version of a *reasons* storyboard that indicates that the teacher should have had the student do more than just copy the proof on the board. There were also responses to the breach versions of the *sequence norm* storyboards that indicated that the teacher's interaction with the student presenter were warranted, even if unexpected. An example of such a response is: "I like how the teacher had Beta erase the steps in the proof instead of adding them in" (27002BC171-Breach). The participant goes on to say: "This emphasizes the need to know what Beta is referring to before using them in a proof" (27002BC171-Breach). This was a response to a breach version of a storyboard in which the teacher interrupts the students transcription of a proof and asks the student to begin the proof again, this time writing the labels in the diagram before they are used in the written statements of the proof. Responses like these—i.e., negative reactions when teachers and students are depicted as following the hypothesized norm, positive reactions when the teacher is depicting as breaching the hypothesized norm—provide evidence that both the breach and control versions of the storyboards are representations of reasonable actions teachers and students could take in classrooms. I revisit this issue in more detail during the report and discussion of the findings from the analysis of the open response data.

Example responses to the follow up questions.

The follow up prompts to the episode and segment appropriateness rating questions asked participants to explain their ratings. The purpose of the follow up prompts was to provide participants with a field where they could anchor their ratings of the teacher's actions to specific aspects of the situation. Such anchoring could take the form of references to specific moments in the storyboard, e.g., "the teacher's reply of 'we know what they meant' is not really ok" (26001BC242), or elaborations that link their ratings to evaluations of a storyboard's people, events, or artifacts, e.g., "all necessary statements need to be written in the proof and not implied" (26001BC183). These are just two illustrative examples. On the whole, responses to the various follow up prompts ranged from general—e.g., "teacher was appropriate"—to specific—e.g., "Skipping over the connection between midpoints and sidelengths." The responses to these fields were coded using the same schemes (described below) as the more general open-response questions but were analyzed separately. This is presented in the section that reports the qualitative data analysis below.

Review of the open response data

Responses from the various open-response fields that participants provided for each storyboard suggest (1) that participants provided a range of responses in the open-responses fields that followed the different storyboards and (2) that the extent to which the responses to the various storyboards are different warrants further investigation. The references to the different aspects of the situations that participants make in the example responses discussed above—e.g., the *little things* that are downplayed by the teacher; the *missing steps* that the teacher fails to correct; the student *copying* a proof onto the board—were coded using a *norm recognition* scheme to identify responses that contain evidence that participants recognized the target norms.

The codes for *norm recognition* were complemented by more general codes for participants' positive or negative evaluations of what takes place in each situation instance represented by the storyboards. The more general codes of positive or negative evaluation were derived from the *attitude system* of the appraisal framework of systemic functional linguistics (Martin & White, 2005). Both classes of codes were used to analyze the open responses to the four types of open response questions (described above).

I describe first the *norm recognition* coding. The description of the *norm recognition* coding includes examples of responses to different questions that illustrate the codes. Following this, I describe the more general coding for positive or negative evaluative comments. After the presentation and exemplification of the coding schemes, I describe the protocol for using the schemes to code the data. The results of the coding are reported next. The chapter concludes with a discussion of the findings from the qualitative analysis.

Coding for Recognition of Semiotic Norms

The storyboards participants viewed and answered questions about were planned to depict teachers breaching or else complying with hypothesized semiotic norms. The *details norm* storyboards represented different ways of breaching the hypothesized routines that secondary teachers have for scrutinizing the details of a geometry proof. The *sequence norm* storyboards represented different ways of breaching the hypothesized expectations that secondary teachers have for how students present work to the class. The purpose of the *recognition* coding scheme is to determine whether there is evidence in an open response that a participant recognized the hypothesized semiotic norms.

The analysis of the responses to the rating questions indicated that participants reacted differently to storyboards where the teacher is depicted as breaching a hypothesized norm when compared to storyboards where the teacher is depicted as complying with a hypothesized norm. The segment-specific rating questions provided a means of linking these differences in ratings more specifically to the segments of the storyboards where the teacher breaches (or does not breach) the hypothesized norm that is targeted by the storyboard. The goal of analyzing the open responses for evidence of *norm recognition* is to provide additional evidence that what was salient for participants in the various storyboards were the departures from the hypothesized semiotic norms, as opposed to some other aspects of the storyboards.

I developed a three-tiered scheme of dichotomous codes to code for recognition of the semiotic norms. The codes at each tier of the scheme are first described generically and then illustrated with norm-specific realizations of the codes with example open response data. Figure 58 is a network representation of each level of the coding.

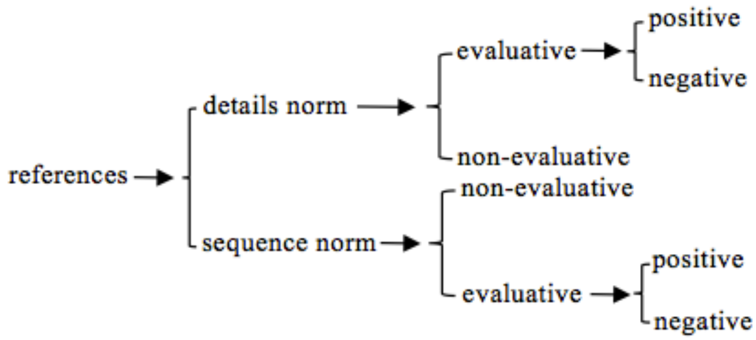

Figure 58: A network representing the three-tiered scheme for recognition coding.

The unit of analysis for the *recognition* scheme was a response. The first tier of the scheme was a dichotomous code (0/1) that indicated whether the response included a direct or indirect *reference* to each hypothesized norm. In the case of the *details* norm, these were direct or indirect references to the level of detail in the proof. In the case of the *sequence* norm, these were direct or indirect references to the student's presentation of the proof to the class. The first tier of the recognition scheme differentiates those responses that contain a general reference to the aspects of a storyboard that are relevant to a norm from those responses that focus on other aspects of a storyboard. Examples and non-examples of the first tier of the recognition coding are illustrated below.

The threshold for the *references* tier of the scheme is low. This is deliberate. The reason for setting a low threshold is to attempt to make operational the ethnomethodological principle of the transparency of the routine. The transparency of the routine is the idea that those aspects of social life that meet our expectations tend to escape our notice. Since the storyboards were scripted in matched breach and control pairs, it is possible to code for general references to the target norms and then investigate whether storyboards that represent breaches of the norm had more of these responses than control storyboards. I develop this idea during the reporting of the open-responses analysis.

The second tier of the scheme differentiated *evaluative references* from *non-evaluative references. Evaluative references* appraise[50] (Martin & White, 2005) the norm-specific aspects of the storyboard that triggered the coding for *references. Non-evaluative references* describe the norm-specific aspects of the storyboard without appraisal. Responses that were coded (1) for *evaluative references* were further distinguished into *positive* and *negative* evaluations.

The purpose of drawing the distinction between non-evaluative (i.e., descriptive) and evaluative references is to make further distinctions between the ways that the details of a proof or the sequence in which a student writes the parts of a proof on the board are noticed by participants. Even if what is routine is generally transparent, it is still the case that participants were asked "what did you see happening in this scenario?" Since something that happened in each scenario is the action that is related to the target norm, it is possible—even in storyboards where the norm was not breached—that participants would reference the details of the proof or the student's presentation of the proof. The second tier of the coding scheme differentiates responses that simply bring up the details of a proof or the student's presentation of the proof from those that evaluate those aspects of the storyboards.

To apply the tiers of recognition coding to the open response data, the responses were blinded with respect to breach/control condition. The purpose of blinding the data with respect to condition was to be able to minimize bias when coding responses for indirect realizations of the *references* tier of the scheme. In terms of recognition coding, the transparency of the routine

[50] The appraisal framework is described below, in the section that describes coding the open responses for positive and negative reactions to the storyboards that are not necessarily focused on a target norm.

would suggest that the storyboards that represent a teacher breaching a hypothesized norm would have more responses that make *references* to what is at stake in a target norm than storyboards in which a norm is not breached. Blinding the data with respect to breach/control condition helps to warrant such analyses. Hypotheses about the frequencies of the different categories of *recognition* coding are stated in the section that reports the results of the qualitative analysis. The next section presents illustrative examples of the different tiers of *recognition* coding, norm-by-norm.

Recognition coding for the details norm

The generic codes for recognition stated above were made operational for the details norm storyboards in terms of references to level of detail in a proof. The first tier of the coding differentiated open responses that reference the level of detail in the proof from open responses that made no such reference. Either: (1) the response provides evidence that the participant commented on how a teacher (in a storyboard) handled the details of a proof, or: (0) there was no such evidence.

Examples of the first level of recognition coding

Responses that reference the level of detail in the proof could do so directly or indirectly. An example of a response that references the level of detail in the proof directly is: "Teacher is adding details to students proof." This is a response to the "What did you see happening in this scenario?" question following storyboard 26013, which is a control dual of a storyboard that depicts a *more details* breach. This response was coded as (1) for *references* because it describes the action of the teacher in the scenario as "adding details" to the student's proof.

Responses could also reference the level of detail in the proof indirectly. Indirect references to the level of detail in the proof include references to their being a "flaw" or an "omission" or a "mistake" in the proof. Such responses were coded as (1) for *references* because the hypothesis underlying the *details* norm is that teachers require certain kinds of details in order for a proof to be acceptable. Apart from the necessary details that are omitted in a proof (*less details* breaches) or the superfluous details that are included in the proof (*more details* breaches), the proofs in the storyboards are representations of typical proofs that are presented in geometry classrooms. It is therefore reasonable to code references to the deficiencies of the proofs as evidence of *references* to the level of detail in the proof. An example of an indirect reference to the level of detail in the proof is: "Omitted the statement to prove sides congruent". This was a response to the "What did you see happening in this scenario?" question following a storyboard that depicted a *less details* breach (26001).

Non-examples of the *references* code make no explicit or indirect reference to the level of detail in the proof. For example, one participant wrote: "A student shared their work on the board and the class was asked to analyze and share their comments about the proof." This was a response to the "What did you see happening in this scenario?" question that followed storyboard 26011, which is the control dual of a storyboard that depicts a *less details* breach. This response was coded (0) for *references* because it does not make any reference to the level of detail in the proof. A different response that was coded (0) for *references* is: "takes a long time to get the proof on the board and then the discussion that ensues takes even longer". This was a response to the "What did you see happening in this scenario?" question that followed storyboard 26001, a storyboard that depicts a *less details* breach.

Examples of non-evaluative references

Responses were coded (1) for *non-evaluative reference* if the response directly or indirectly referenced the target norm without evaluation. An example of a response that was coded (1) for *non-evaluative reference* is: "Teacher is adding details to student proof".

(26003BC141-Breach). This response, also cited above, was coded as a *non-evaluative reference* because "adding details" is a description of an action that the teacher takes in the scenario. Because of the explicit reference to "details", it was coded as (1) for *references.* But it is not clear, from the response alone, whether the participant is saying that this is a good or bad move on the teacher's part. One could argue that the act of "adding details" to a proof is inherently positive, however there are alternative phrasings that are more clearly positive, such as describing what the teacher does as: "…providing the correction" (26003BC191-Breach) or as telling a student: "…what they are missing." (26003BC101-Breach). Referring to what the teacher does as "adding details" is a means of describing what takes place in the situation in a non-evaluative way.

Examples of evaluative references

Responses were coded (1) for *evaluative references* if the response positively or negative appraised the reference to the target norm. An example of a *negative evaluative reference* is: "allowed incomplete proof to pass despite student correction of incomplete proof." This is a response to the "How appropriate was the teacher's review of the proof?" question following storyboard 26001, a storyboard that depicts a *less details* breach. In this response, the participant specifically references the "incomplete proof" and suggests that the teacher should not have accepted the proof without modification. Another example of a *negative evaluative reference* is: "teacher does not validate student's correct and very clear criticism of the proof. She falls into the trap of skipping small but meaningful details." This was a response to the follow up prompt of the *segment of interest* appropriateness rating for storyboard 26002, a storyboard that depicts a *less details* breach. This is a *negative evaluative reference* because of the negative judgment of the teacher's handling of the details of the proof—i.e., the teacher skipped "small but meaningful details" of the proof.

An example of an *evaluative reference* following a *more details* storyboard is: "The teacher is asking for a degree of thoroughness that the typical student is not going to be interest[ed] in achieving" (review-of-proof rating question, *more details* breach storyboard). This response points specifically to the teacher asking for an unreasonable degree of thoroughness. The "degree of thoroughness" is the reference to the level of detail in the proof, while the reference to the "typical student" that will have "no interest" in achieving that degree of thoroughness is an indication that such a level of detail is unnecessary. Hence, this was response was coded as a *negative evaluative reference*. A second response that illustrates *evaluative references* for the *more details* storyboards is: "The rays intersected by definition. We don't need a theorem to justify it" (response to segment-of-interest rating question, *more details* breach storyboard). In this response, the statement that the "rays intersect by definition" is an indication that the information that the teacher is requesting is tacitly conveyed by the diagram. The "we don't need a theorem" statement is an indication that what the teacher is requiring is unnecessary.

Recognition coding for the sequence norm

The first tier of recognition coding for the sequence norm storyboards distinguished responses that contained a reference to the student presentation of the proof from those that contained no such references. The references to the student presentation of the proof could be direct or indirect. Examples of each kind of reference are illustrated below.

Examples of the first level of recognition coding

Responses could contain direct or indirect *references* to the student presentation of the proof. An example of a response that contains a direct reference is: "A teacher was guiding discussion of a problem that a student put on the board." (27001BC111-Control). In this response, there is a direct reference to a problem that a student put on the board. Another example of a direct reference to the student presentation is: "Beta was invited to show his proof on the board and he did so." (27002BC141-Control). In this example, the student, Beta, is described as showing his proof in the board. Responses that directly reference the student demonstrating, or presenting, or showing work on the board were coded (1) for *references.*

Responses could also be coded (1) for *references* if they indirectly referenced the student presentation of the proof. An example of such a response is: "Students are sharing proofs in a geometry class." (27005BC191-Control). This response describes the activity taking place in the classroom as "students sharing proofs in geometry class". Since the "sharing" that takes place in the storyboard is a student at the board presenting a proof to the class, this response was coded (1) for *references.* Another example of a response that indirectly references the student presentation of the proof to the class is: "The teacher had a student go over a homework problem" (27002BC31-Breach). This response was coded as a reference to the student presentation of the proof because it describes what the student did as "going over a homework problem" and in the storyboard, the teacher calls a student to the board to go over a problem from the homework. A third example of a response that indirectly references the student presentation of the proof to the class is: "student forgot a process in the proof, teacher wanted her to correct it" (27004BC241-Breach). This was coded as a reference to the student presentation because it indicates a specific moment in the student presentation of the proof, even though it does not explicitly establish that the student was presenting the proof to the class at the moment the student forgot the "process".

Responses that indirectly reference the student presenting a proof to the class are distinct from responses that contained no such reference. An example of a response that does not reference the student presentation of the proof is: “the teacher is validating how differnt meathods are acceptable (*sic*)” (27001BC181-Control). This response was coded (0) for *references* because the focus of the response is the teacher’s validation of the different methods for doing the proof that were discussed during the storyboard. The conversation about the different methods takes place in parallel to the student presentation of the proof. It is not clear if the response means to include the student presentation on the board as one of the different methods that the teacher validates. Because the response indicates the teacher (not the student) as actor and references different methods (as opposed to a focus on the proof that was presented), this response was coded (0) for *references.* Another example of a response that was coded (0) for *references* is: “Also good math but much better chemistry between teacher and students” (27001BC121-Control). In this example, there is a reference to “good math” and the “better chemistry” between the teacher and students. These could be references to the way that the teacher handles the student presentation of the proof to the class, but there are other aspects of the storyboard that would also fall under “good math” and “better chemistry”. This response was therefore coded (0) for *references.*

Examples of non-evaluative references

Responses that were coded (1) for *non-evaluative references* if the response used descriptive—rather than evaluative—language to reference the student presentation of the proof. An example of a *non-evaluative reference* is: “Student putting a proof on the board - other students having a conversation” (27002BC131 - Control).” The response is non-evaluative because it provides a descriptive account of the activity of presenting the proof to the class. A second example of a *non-evaluative reference* to the *sequence norm* is: “The teacher asked for a volunteer to come to the board to complete a proof from the homework.” (27003BC181-Control). This response is a *non-evaluative reference* to the student presentation of the proof because it reports what the teacher asked the student to do without appraising the doer (the student) or what was done (completing a homework proof). A third example of a *non-evaluative reference* is: “Student showing work on the board” (27005BC71-Control). Like the other examples of *non-evaluative references*, this response describes the student presentation of the proof without appraising it.

Examples of evaluative references

Responses that were coded (1) for *evaluative references* positively or negatively appraised the student presentation of the proof. An example of a response that contains an *evaluative reference* is: “More students are engaged, but I don't like the fact that the teacher put the student on the spot to try to change her proof while up at the board. The student was prepared to label it her way**,** why did the teacher have to analyze the students comprhension (*sic)* in that way?” (27001BC131). This was a response to the “What did you see happening question?” for a *sequence (labels)* breach storyboard. This is an example of a *negative evaluative reference* to the student presentation of the proof. The response negatively appraises the fact that the student was “put on the spot” and questions why the teacher intervened when “the student was prepared to label it her way”. This response uses negative evaluative language to reference the student presentation of the proof and was therefore coded as a *negative evaluative reference.* Another example of an *evaluative reference* is to the student presentation of the proof is: “I think the teacher created unnecessary work for the student. They labeled their paper with the correct labels and overlooked them on the board. Having them erase steps and start over is ridiculous.”

(27002BC151) This response is also an instance of a *negative evaluative reference* to the student presentation. In this case, the response indicates that the teacher created "unnecessary work" that was "ridiculous" because the student had already labeled the figure correctly on the homework paper.

There were also responses that were coded as *positive evaluative references.* An example of a *positive evaluative reference* is: "Beta was invited to show his proof on the board and he did so. Teacher monitored the room and kept the focus on the proof at hand, tabling related discussion appropriately" (27002BC141-Control). This response does not appraise Beta's presentation of the proof, however, it is still an instance of a *positive evaluative reference* to the student's presentation because of the indication that the teacher acted "appropriately" when the conversation about Beta's proof was "tabled". This is a reference to the teacher telling the students that ask a question while Beta is in the process of transcribing the proof to hold their comments until Beta has finished. The fact that it was appropriate for the teacher to take this action is the reason that this response was coded as a *positive evaluative reference* to the presentation.

Another example of an *positive evaluative reference* is: "This played out just like I'd want it to, I think. Teacher asking the student to put something on the diagram they forgot, then asking questions and helping with the flow of the conversation without dictating it..." (27001BC131-Control). The fact that the scenario played out "just like" the participant would want it to is a *positive evaluative reference* to the work that the teacher does to support the student who is presenting the proof to the class. A third example of a *positive evaluative reference* to the student presentation is: "I think the teacher did a good job of explaining why the markings should not be on the diagram and even though they had it correct in the homework, the student still needed to do it correct on the board because the rest of the students would be looking at it." (27003BC102-Breach). In this example, the teacher is positively appraised for correcting the student while the student is in the process of presenting the proof at the board. The responses indicates that the teacher did a "good job of explaining" to the student that the use of markings at the proof at the board was separate from what the student had already completed for homework.

Summary of the recognition coding scheme

The coding for recognition proceeded via the tiers defined above. The design of the recognition coding scheme allowed me to code the open response data without knowing whether a response was associated with a breach or control storyboard. The definitions of the codes for *references*, *non-evaluative references*, and *positive* or *negative evaluative references* did not depend on the features of the storyboards that were specific to their breach or control realizations. The tiered coding for recognition is a means of making the expected transparency of the routine operational for analyzing the coded qualitative data. In the section (below) that reports the qualitative findings, I state hypotheses about the relative frequencies of the different categories of recognition coding. The next section describes the scheme I developed to code for more general positive or negative reactions to the storyboards.

Coding for Positive and Negative Reactions to the Storyboards

The analysis of open response data reported above aimed to identify responses where participants make a reference to a semiotic norm. Such specific responses provide one kind of evidence that participants noticed whether the teacher breached the target norm in each storyboard. But there are other ways that participants could signal that something unexpected happened in a storyboard. Chapter 4 describes the design of the study as a virtual breaching experiment with control. Using the descriptions of classic breaching experiments as a guide, it is possible that participants reacted in a non-specific negative way to the storyboards that depicted a breach of the norm. Garfinkel (1963) indicates anxiety, anger, and bewilderment—as well as other negative emotions—as feelings that people experienced when confronted with a planned departure from the expected social order. Participants in situations that unfold in ways that depart from the (tacit) expectations of the shared social order might not be able to specifically describe what is causing them to feel anxious or angry, yet nevertheless, the departure from what routinely occurs can evoke these emotions. In the case of the virtual breaching experiment with control, this suggests that, in response to the storyboards that depicted a teacher breaching a norm, participants could signal that they noticed the breach by conveying negative feelings about the situation in the open responses.

Since the storyboards that breach the details norm, by hypothesis, are designed to represent specific departures from the expected routines for scrutinizing the details of a proof when checking proofs in geometry (*details norm*) or facilitating student presentations of proofs at the board (*sequence norm*), another way of measuring the extent to which participants reacted to the breach of the norm is to analyze the open responses for expressions of negative feelings about the storyboards. The general hypothesis would be that the open responses to storyboards that breach the norm have more expressions of negative feeling than the open responses to storyboards in which the norm is not breached. The analysis reported below draws on the *appraisal system* of systemic functional linguistics to develop a scheme that makes the notion of "negative feeling" operational for coding the open response data (Aaron, Erickson, Dimmel, & Herbst, 2013)

The next section summarizes the *attitude* system of the appraisal framework and describes how the distinctions between different ways of construing attitude informed the categories in the coding scheme. Following this, I present the categories in the scheme with examples of the codes. Once the scheme and examples of codes are presented, I report on the analysis of the coded data.

An appraisal-based scheme for coding open response data

The central challenge for analyzing the open responses for evidence of more general indications that participants are reacting to the breach of the norm is classifying when a response indicates that a participant responded positively or negatively to a storyboard. Since the storyboards are episodes of geometry teaching during which—even in instances where the teacher breaches the norm—the teacher behaves in ways that are recognizably professional—i.e., the teacher does not take any extreme or irrational departures from what is expected, for instance by verbally threatening a student in the class—and the situation itself is identifiable as an episode of geometry instruction, it is possible that the negative reactions to the situation could be relatively muted. Any scheme for coding whether a response indicates a positive, negative, or no (e.g., purely descriptive) reaction toward the storyboard that is the target of the response would need to be able to capture potentially subtle variations in response. The appraisal framework (Martin & White, 2005; Read & Carroll, 2012) of systemic functional linguistics provides a model of how people convey feelings through choices in language. This framework has the necessary degree of sensitivity for coding these potentially subtle differences.

In *The Language of Evaluation,* Martin and White (2005) define *attitude* as a system of language choices through which people convey positive or negative feelings. They identify three classes of attitude: *affect, judgment,* and *appreciation.* Coarsely, these different categories of attitude correspond to accounts of personal feeling (affect), evaluations of people and their deeds (judgment), and qualitative statements about events or things in the world appreciation). While these different ways of feeling have different lexical and grammatical realizations, they are conceptually linked in Martin and White's (2005) model. They define judgment and appreciation as *institutionalizations* of affect.

In the case of judgment, affect is institutionalized through social mores, ethics, or other shared codes of acceptable social conduct (Martin & White, 2005). The system of judgment thereby provides a set of choices for appraising people and their deeds. "He is an outstanding teacher" is a statement that conveys a positive judgment about a person who is fulfilling the role of teacher.

In the case of appreciation, affect is institutionalized through a shared sense of the qualities that things in the world ought to have in a given context (Martin & White, 2005). "The rainfall was spectacular" could be an institutionalization of the gratitude or joy that a person in a drought-ravaged area might experience during a thunderstorm. "His explanation was outstanding" could be an institutionalization of the pleasure one can experience (Feynman, 1999) when presented with a clear, complete account of why something occurs.

The distinction between affect, judgment, and appreciation provides a starting point for analyzing the open response data for evidence of positive or negative feelings. The appraisal framework that is developed by Martin and White in the *Language of Evaluation* (2005) describes additional levels of delicacy for each of these systems. They distinguish different types of judgment—e.g., capacity, veracity, tenacity, propriety, usuality—and give examples of linguistic markers through which these different judgments are realized, e.g., *tenacity* refers to judgments of a person's determination, as in "he never quits"; *capacity* refers to judgments of person's abilities, as in "he is a skilled debater". Appreciation has levels of delicacy that, for example, differentiate *composition*—as in: "the explanation was clear"—from *balance*—as in: "the porch was lopsided". In a larger corpus, these more fine-grained levels of delicacy might suggest different ways that participants are reacting to the storyboards they viewed—e.g., the breach and control version of a storyboard might differ in the number of negative judgments of a teacher's capacity, but maybe they exhibited comparable numbers of positive judgments of student tenacity. Given the relatively small size of the corpus of open response data for this study, however, these additional levels of delicacy are not likely to have sufficient realizations to warrant this kind of finer-grained analysis.

In lieu of coding the corpus for specific differences within kinds of attitudes—e.g., judgments of capacity, appreciations of composition—I developed a coding scheme that draws on the appraisal system to analyze the corpus for evidence of positive and negative attitudes in participants' open responses. Since each open response is authored by a single person, the *appraiser* in each response is the participant. The categories of interest then are the *appraisal*—in the case of this coding, the kinds of attitude that a participant is conveying in the response—what is being *appraised*—that is, the target of the appraisal—and the *polarity*—positive or negative attitude toward—of the appraisal.

I represented the scheme that I developed in a systemic network (Halliday & Matthiessen, 2004), shown below in Figure 59.

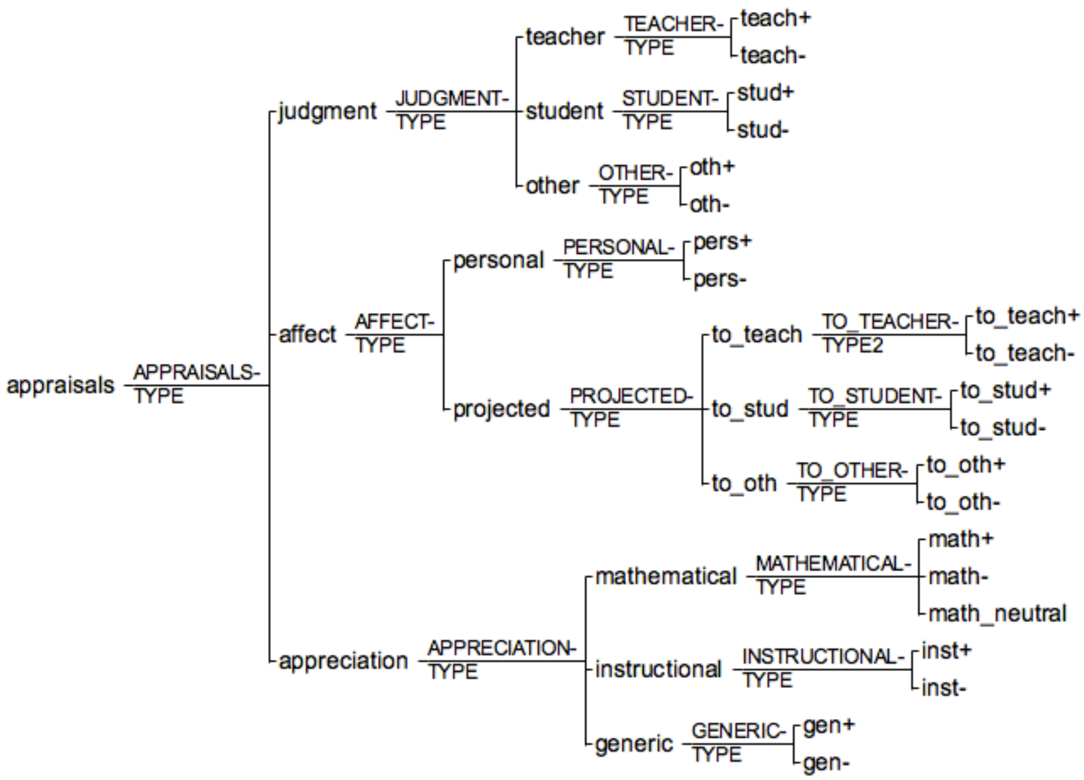

Figure 59: A system network representing the choices in the scheme that was used to code the open response data. System network produced using UAM CorpusTool version 3.0.

In this network, all of the choices are exclusive–or choices, meaning that any attitude is coded into one and only one of the terminal categories in the scheme. The entry condition of the network is the type of attitude—judgment, affect, or appreciation—that is evident in the response. The next choices in the scheme are for the target of the attitude. While there are potentially a large number of targets for any statement of attitude, it is reasonable to partition this space into a set of choices for targets that are relevant for this study. In the case of judgment, for example, while a participant certainly could make judgments about any number of people, the judgments that are expected and relevant to this study are judgments about the teacher and students in the situation. Thus, under the category *judgment*, the choices for targets of judgment are *teacher, student,* and *other.* The *other* category allows the scheme to be exhaustive while also providing a means to examine the corpus for evidence of storyboards that evoked a critical mass of unexpected judgments. If a storyboard featured an unlikely preponderance of negative judgments of a student's parents, for example, such judgments would be coded as *other* and would be reviewable by searching the corpus for instances of *other* judgments.

Coding for Judgment: lexical and grammatical realizations

An example of a response that conveys a positive judgment of the teacher is: "teacher is guiding students effectively" (response 26004BC133). This is an open response that a participant provided to the targeted rating question following the control version of storyboard 26004—the *more details* storyboard about the angle bisectors of consecutive angles in a parallelogram. The teacher is the target of the attitude, while "effectively" is a positive evaluation of the action—guiding—that the teacher is described as doing in the scenario. Since this is about the quality of how a person (i.e., the teacher) performs an action, it is coded as a positive judgment.

The example discussed above is straightforward because it contains an explicit lexical marker or evaluation—the adverb *effectively*. The presence of such explicit lexical markers provides a condition for distinguishing evaluation from description. There are other responses, however, that still seem to convey an evaluation of the teacher (or student) even in the absence of such explicit markers. The challenge in such responses is to distinguish descriptive from evaluative accounts of what happens in the scenario. An example of a response whose evaluative or descriptive quality is more difficult to gauge is response 26001BC214: "[The] Teacher *should have explained* that othe[r] student's proof was correct too…" (emphasis added) A second, even more difficult example is response 26004BC104: "The instructor listened to the student question and the student response, *then challenged them to think more* (emphasis added)".

Neither of these examples has specific lexical markers of positive or negative evaluation of the actions that the teacher does. Yet both responses are examples of response-types where there is evidence of evaluation. The first response is an example of a *grammatical* realization of a judgment (Herbst, Aaron, Dimmel, & Erickson, 2013; Herbst & Dimmel, 2011; Martin & White, 2005). The response conveys a negative evaluation of what happens by describing an alternative that should have happened instead (Chieu & Herbst, 2011). The use of the modal (obligatory) *should* signals the unsatisfactory status of what *did* occur. Instances where participants offered alternatives that the teacher in the storyboard *should have done,* or else described alternative actions that they *would have done* or that even stated alternatives that one *could have* or *might have* done in the situation were coded as negative judgments of the teacher. Responses where participants described the teacher's actions as something that they *would do* or that all teachers *should do* in their classroom were coded as positive judgments of the teacher.

While responses like the first difficult example do not contain lexical markers of positive or negative evaluation, the modality statements they contain indicate that they are grammatical realizations of judgment. The second example listed above—response 26004BC104—is more difficult still because, at first glance, it appears that it is a purely descriptive statement of what the participant witnessed the teacher do during the storyboard: "The instructor listened to the student question and the student response, *then challenged them to think more*" (emphasis added). But upon closer inspection, this account of what the participant saw taking place in the storyboard seems different from more purely descriptive accounts of the same part of the same storyboard, such as: "Oral review of the progress of the proof" (26004BC164). Where one participant sees an "oral review of the progress of the proof", a different participant sees a teacher who is *challenging* the students to *think more*. Response 26004BC104 imputes an action—*challenging*—to the teacher that effects an undoubtedly positive state of affairs in the classroom—*more student thinking*. Furthermore, this is the only response in the corpus that ascribes "challenging the students to think more" as the action the teacher takes in the situation. Coupled with the fact that there are other responses to the same prompt for the same storyboard that are more basically descriptive than this response, I decided to code it as an instance of positive judgment of the teacher.

Response 26004BC104 suggests a second general principal for coding grammatically realized judgments of attitudes: If the account of events casts the teacher as playing a principal role in bringing about a positive or negative state of affairs, the response is coded as a judgment of the teacher. Another example that is coded as a judgment of the teacher in accordance with this heuristic is response 26002BC133: "The teacher gave a very broad question *that may not have instigated the students to think specifically*." Although "giving a very broad question" could be descriptive, this action by the teacher may not have been the most productive for helping the students to think about the task at hand. This response was therefore coded as a negative judgment of the teacher. Throughout all open responses in the corpus, participants made 504 judgments. Of these, 416 were judgments of the teacher, 86 were judgments of students, and 2 were other judgments. More detailed analyses of the results of the attitude coding are reported below, following the elaboration of the other codes in the scheme.

Coding affect: personal and projected feelings

In addition to making evaluative statements about the teacher and students in the storyboards, there was also the potential that participants would express more direct statements of feelings in their open responses. The attitude category of *affect* provided a starting point for coding these responses. I distinguished two general classes of affective response. A *personal* *affect* is a statement participants make about their own personal emotional state, e.g., "I am bored just watching the slides." A *projected affect* is a statement that participants make about the emotional states of either the teacher or students in the scenario, e.g., "students feel comfortable" As was the case for judgment, an *other* category was added as a choice under *projected to* in the interest of making the scheme exhaustive. Also, like judgment, each choice for the type and target of a statement of affect could be positive or negative. Throughout the corpus, there were a total of 49 instances of participants indicating affect in their responses. Of these, 41 were projected affects—almost all of which were projected to students—and 8 were personal affects.

Coding Appreciation: mathematical and instructional evaluations

The final type of attitude that I coded the open responses for is *appreciation.* Since the targets of appreciations are goods of production or things in the world, in the case of the current study I distinguished mathematical, from instructional, from generic appreciations. Here, as before, the inclusion of the *generic* category ensures the exhaustiveness of the scheme. Mathematical appreciations include statements about a proof, the statements in a proof, the diagram that accompanies a proof, or the overall quality of the proof as a mathematical good. Other statements that appraise the mathematical aspects of the situation are also included in this category.

An example of a mathematical appreciation is "the incorrect proof", where *incorrect* is a negative-valued appreciation (perhaps of the proof's composition, or balance). An example of a more general mathematical appreciation is: "different solutions to a problem are a good thing". The mathematical target of this appreciation is the alternative method for completing a proof that is shared by a student in 26001. The participant's response conveys an attitude that such alternative solutions are “good things”. Still another example of mathematical appreciation of the proof is the response: “the math proof was not accurate.”

One difference between *mathematical appreciations* and other categories in the coding scheme is the possibility that a mathematical appreciation could be neutral—that is, neither positive nor negative—as is the case with the statement: “one small step in the proof”. While *small* is a way of appreciating the step in the proof, it is not clear that this is necessarily a positive or negative statement about the step (or the proof). It is rather a way of evaluating the role that the step plays in the proof. The fact that it is a small step, however, does not mean that it is minor or unimportant.

Another instance of a neutral mathematical appreciation is that of describing a proof as “basic”. While such a description is an appreciation of the proof—in this case, it is an appreciation of the proof’s complexity—it is not clear that such an appreciation has a positive or negative value. In some contexts, *basic* could be fundamental, simple, general, and one could make an argument that it would therefore be positive-valued. At the same time, basic could mean rudimentary, remedial, unnecessary, in which case one could argue that this is a negative appreciation of the proof. In order to make such a determination, the responses would need to provide other markers of the participant’s attitude toward the proof, e.g., “wasting time on a basic proof” would, perhaps, provide sufficient evidence that “basic” in this context is meant to be a negative statement of the proof’s complexity (and therefore worthiness). In the absence of such additional information, these and other ambiguous mathematical appreciations were coded as having a *neutral* value.

Instructional appreciations include positive and negative attitudes that participants express toward the events that the teacher brings about in the scenario. “It was a good discussion", for example, is a positive instructional appreciation, on the grounds that discussion is a product of the activity that takes place in classrooms. Another example of a positive instructional appreciation is: "the teacher's explanation was correct." In this example, the explanation that the teacher provides is positively appreciated. Since it is the explanation—rather than the teacher-as-explainer—that is the target of the appraisal, this is an appreciation, rather than a judgment. The attitudes in these examples are conveyed so that events in the situations—i.e., the discussion itself, the teacher's explanation—are the targets of the appreciation. One could imagine a related attitude where the teacher would be the target of the attitude expression, as in: "The teacher did a good job with the discussion", or "the teacher is a good explainer." What choices in the type of attitude—affect, judgment, or appreciation—suggest about participants' reactions to the storyboards will be considered during the discussion of the qualitative analysis, below.

The *generic* category of appreciation captures other aspects of the situation—besides the mathematical or instructional—that participants could appraise. This category also provides a way of coding appreciations whose targets are ambiguous, as in: "it was fine". This is an example of an open response that is included as a follow up to the targeted distracter rating question that accompanied storyboard 26001. Since the rating question was about the teacher's actions, it could be warranted to conclude that the "it" refers to the teacher's actions. In which case this response would be coded as a positive instructional appreciation. There are other possibilities for the referent of the "it" however, such as the alternate method of completing the proof that a student provides. In which case this response would be coded as a positive mathematical appreciation. Given this ambiguity, coding "it was fine" as a generic positive appreciation seems preferable.

In the next section, I describe the method I used to apply the appraisal scheme to open response data. Following this, I report tallies of the results of the attitude coding for the entire corpus, and by breach/control condition. The final section states hypotheses about the frequencies of different kinds of attitudes in open responses to breach and control storyboards, and reports the results of statistical testing of these hypotheses.

Curating and coding the open response data

The raw open response data[51] for each open response field of each storyboard was downloaded to a spreadsheet. From the raw data, I created a curated corpus that blinded whether a response was associated to a breach or control version of a storyboard. Both the storyboard that a response accompanied and also the prompt that it was linked to were preserved as metadata in the name of each blinded open response. To create the blinded data file, for each response type—"1" for the general open ended question; "2" for the follow up to the general rating question; "3" for the follow up to the targeted rating question; "4" for the follow up to the targeted distracter rating questions—I merged the responses from the breach version of the storyboard with the responses from the control version of the storyboard, listing the breach versions of the responses first. This merged data was a single column of open responses, stratified by response type.

For each storyboard, the number of responses for each response type is equal to the sum of the participants that completed the breach and control versions of the storyboard. For example, there were 29 entries for each type of response in the merged file for storyboard 26001, because 16 participants completed the breach version of the storyboard and 13 participants completed the control version of the storyboard. I used a random sequence generator to create four different

[51] This data is securely stored on Lesson*Sketch* server, in accordance with the data management protocol for project Themat (principal investigator is Pat Herbst).

random orderings of the number of responses to each question of each storyboard. I then used these random sequences to index the original responses. Sorting—from 1 to 29—the responses to each question type by these sequences created a mixed list of breach/control responses. I then copied the original and randomly indexed response numbers to a separate file—i.e., I created a key to unblind the file—and deleted the original index from the now blinded corpus of open response data.

I assigned IDs to each open response using the following format: *storyboard_nameBC[random index entry][response type]*.Response 26003BC102, for example, is a response that is associated to either the breach or control version of storyboard 26003. In the blinded data file, this is the 10th response to the second open response question. This procedure ensured that as the open response data were analyzed, the coders did not know whether they were coding a response that was associated to a breach or control version of a storyboard. Blinding the data in this way was important to the integrity of the analysis because of the issues of distinguishing evaluation from description that are described above. Since a primary hypothesis underlying the design of the virtual breaching experiment with control instrument is that people could react in vaguely negative ways when they encounter a breach of the expected social order, blinding the data helped to curtail the potential for bias to enter the coding of the open response data.

The blinded corpus was imported to the UAM CorpusTool environment (O'Donnell, 2008) as a collection of .txt files. The .txt files were stored in storyboard-specific sub-corpora and analyzed storyboard-by-storyboard. Each storyboard specific sub-corpora consisted of the blinded responses to a breach/control pair—e.g., the 26001 subcorpus contained the the blinded responses to the breach (26001) and control (26011) realizations of the first storyboard that targeted the details norm. I imported the schemes to code each response for recognition of semiotic norms and general reactions to the situation that are described above. The schemes were created in separate coding layers and the coding proceeded in waves. For each storyboard-specific sub-corpus, I coded first for recognition of semiotic norms, using the norm recognition scheme described above. Following the coding for recognition, I coded each response for evidence that participants repaired the situation, using the attitude scheme described above. The implementation of the coding layers was independent, meaning that, when coding for general appraisals, I did not know how a response had been coded for norm recognition. The mechanics of the coding involved using CorpusTool's text selector function to highlight a segment of a response, then selecting—by successively clicking—the categories from a scheme that applied to that segment or entire response.

Reliability of the Recognition and Appraisal Coding Schemes

The *recognition* and *appraisal* schemes were tested for reliability by comparing coded responses of two independent coders[52]. The coders applied each scheme to 100 randomly selected texts in the corpus—25 of each of the 4 response types, roughly 10% of the total number of responses. The *kappa* statistics for the *recognition* scheme are .89 and .76 for the *details* and *sequence* norms recognition coding; the *kappa statistics* for the *appraisal coding* for which there were sufficient instances of the codes to warrant the statistics are .79 for *negative judgments* of the teacher; .49 for *positive judgments* of the teacher; .77 for *negative mathematical* appreciations; .49 for *positive mathematical* appreciations. These *kappa* scores indicate moderate (.49), high (.76, .77, .79), and very high (.89) agreement between the coders.

[52] I acknowledge the support of Nicolas Boileau for his assistance with the reliability study of these schemes.

Analysis of the Coded Open Response Data

I report here the results of the coding of the open response data. I report first the results of the coding for *norm recognition*, by target norm and breach/control condition for each of the four open response questions. I next report the results for the more general appraisal coding. I discuss the results as I report them, drawing on examples of coded responses to illustrate the findings.

Results of recognition coding

I described above the ethnomethodological principle of the transparency of routine—that those aspects of a situation that are expected to happen generally escape our notice. The matched breach/control design of the storyboards that were used as probes in the study provides a way to use this principle to investigate the coded open responses data. After the data was coded using the *recognition* scheme, the data was unblinded with respect to the breach/control condition of the storyboard. I then compared the results of the recognition coding at each tier of the scheme across the breach/control conditions of the storyboards. The goal of the comparisons is to investigate whether the open response data indicates that participants reacted to the breaches of the hypothesized norms. The results of the comparisons are reported below, by tier of the *recognition* coding scheme. Where possible, I used chi-square to test for a significant association between storyboard condition and the categories in the *recognition* scheme. In those instances when chi-square could not be used[53], I tested for significant association using Fisher's exact test.

At the first tier of the *recognition* scheme, I hypothesized that responses to the "What did you see happening in this scenario?" question would be more likely to contain references to the target norm in storyboards that represent a breach of hypothesized norm than in storyboards that represent the teacher as not breaching any norms. At the second tier of the scheme, I

---

[53] I followed the standard requirement that the expected value of each cell in a 2x2 contingency table needs to be greater than or equal to 5 in order for the chi-square test to be valid.

hypothesized that, within the set of responses that contained a reference to a target norm, there would be more evaluative than non-evaluative references in the case of storyboards that represent a breach of the norm. Table 34-Table 37 report tallies of responses that were coded for different kinds of references, stratified by norm and condition.

In the tables, the numbers in each cell are the total number of responses to the "What did you see happening in this scenario?" question that contained the type of *reference* that is specified in the column header. The categories at each level in the scheme were mutually exclusive (i.e., a response either contains a reference to the target norm or it doesn't). The results of each table are discussed below the table.

Table 34: Counts of responses that contained a reference to the target norm for the first open response question, by norm and condition.

Norm/Condition	**No Reference**	**Reference**	**Total**
Details/Breach Storyboards	33	25	58
Details/Control Storyboards	38	18	56
Sequence/Breach Storyboards	11	77	88
Sequence/Control Storyboards	29	61	90

Table 34 indicates that, for both the *details* and *sequence* norm storyboards, there were more responses that contained *references* to the target norm in the breach condition than in the control condition. In the case of the *sequence norm*, a chi-square test of association indicates there is a significant relationship between the breach/control conditions and the categories of *no* *reference* and *reference* ($\chi^2 = 8.83$, $p = .003$)[54]. There is no significance relationship in the case of the *details norm.*

One possible reason that there were a relatively large number of responses to the *details* norm storyboards that make a direct or indirect reference to the norm in storyboards where the norm is not breached could be the explicit way that the teacher deals with the details of the proof in 3 of the 4 storyboards that target the *details norm*. In the case of each of the *less details* storyboards, in both the breach and control conditions, the level of detail in the proof is an issue that explicitly enters the discourse of the classroom. In the case of the control versions of these storyboards, the comments about the level of detail in the proof spur the teacher to act in ways that are designed to be compliant with the hypothesized *details norm.* In the case of the breach versions of these storyboards, the comments about the level of detail in the proof spur the teacher to act in ways that are designed to represent departures from the hypothesized *details norm.* Thus, across both conditions for 3 out of the 4 *details norm* storyboards, the amount of detail in the proof is anchored to actions that the teacher takes.

The fourth *details norm* breach-control storyboard pair does not fit this pattern. In the control version of this storyboard, the teacher acts in ways that are designed to comply with the hypothesized norm: the teacher shows the angle bisectors of a quadrilateral intersecting without problematizing the existence of a point of intersection for the rays. In the breach version of this storyboard, the teacher asks the students in the class to provide a warrant that the two rays do, in fact, intersect at a point. Because the control version of this storyboard depicts the teacher as

---

[54] I created separate 2x2 contingency tables for the *details* and *sequence norms,* because I only coded each response for *recognition* of the target norm. An area for further study would be to investigate whether there are references to the details of the proof in the *sequence norm* storyboards and whether there are references to the student presentation of the proof in the *details norm* storyboards.

complying with the hypothesized norm without explicitly stating anything about the details of the proof that are being provided by the diagram, the responses to this storyboard provide a best case scenario that the responses to the "What did you see happening in this scenario?" question would break according to the expected transparency of the routine hypothesis stated above. And in fact, this is the case.

For this storyboard set, there were 13 responses to the breach version of the storyboard and 15 responses to the control version of the storyboard. Table 35 reports the results of the first tier of the references coding for this storyboard, stratified by condition.

Table 35: Counts of responses that contained a reference to the target norm for the first open response question, by storyboard and condition.

Storyboard (Condition)	**No Reference**	**Reference**	**Total**
26004 (Breach)	8	5	13
26014 (Control)	15	0	15

The expected values of the *references* cells are less than 5, meaning it is not appropriate to use chi-square to test for a significant association. Using Fisher's exact test, however, yields a *p*-value (0.013) that indicates that the association indicated in Table 35 is significant.

That the control version of storyboard 26004 represents a teacher complying with the target norm in more tacit ways than in the other *details norm* storyboard helps to shed light on the fact that there was a significant association at the first tier of the *recognition* coding for the *sequence norm* storyboards yet no such association for the *details norm* storyboards. In the control versions of the *sequence norm* storyboards—in which the student transcribes a proof in an order that makes the transcription convenient but that renders the proof logically incoherent as it is being written—the student is depicted as complying with the hypothesized norm in ways that could be considered tacit. The student presentation of the proof was designed to represent the hypothesized-to-be-natural transcription practices that students use when writing proofs on the board in geometry classrooms. In control versions of the *sequence norm* storyboards, it is possible that the student's presentation of the proof was referenced in the open responses less frequently because it faded into the background.

The results of the first tier of recognition coding indicate that, in storyboards where the teacher and students are depicted as complying with the hypothesized semiotic norm in ways that are more tacit, there were more references to the target norm in breach storyboards than in control storyboards. The second tier of the recognition coding distinguishes those responses that make evaluative references to the details of a proof or the student presentation of a proof from those that do not make evaluative references. In the case of storyboards that depict a teacher breaching a hypothesized nom, I hypothesized that there would be more evaluative references than non-evaluative references. The results of the second tier of the *recognition* coding are reported in Table 36.

Table 36: Counts of responses that contained evaluative references to the target norm for the first open response question, by norm and condition.

Norm/Condition	Evaluative Reference	Non-evaluative Reference	Total
Details/Breach Storyboards	24	1	25
Details/Control Storyboards	16	2	18
Sequence/Breach Storyboards	49	28	77
Sequence/Control Storyboards	22	39	61

For the *details* and *sequence* norm storyboards, there were more evaluative than non-evaluative references to the target norm in storyboards that represented breaches of the target norm. In the case of the *sequence norm,* a chi-square test of association indicates that there is a significant relationship between breach/control condition and whether there is an evaluative or non-evaluative response (χ^2= 10.36, p =.001). This result is consistent with the differences between tacit versus explicit ways of complying with the norm that were described during the reporting of the first tier of the recognition scheme above.

The final level of the recognition coding distinguishes *positive* from *negative* evaluative references. These results are reported in Table 37.

Table 37: Counts of responses that contained positive and negative evaluative references to the target norm for the first open response question, by norm and condition.

Norm/Condition	**Positive Evaluation**	**Negative Evaluation**	**Total**
Details/Breach Storyboards	8	16	24
Details/Control Storyboards	12	4	16
Sequence/Breach Storyboards	7	42	49
Sequence/Control Storyboards	10	12	22

The results reported in Table 37 indicate that there were more responses that contained negative evaluative references than positive evaluative references following storyboards that represented a breach of a target norm. The results also indicate that there were more responses that contained positive evaluative references than negative evaluative references following storyboards that represented compliance with a target norm. In both cases, a chi-square test of association indicates a significant relationship between breach/control condition and the polarity of the evaluative reference (χ^2= 6.67, p =.009 for the *details* storyboards; χ^2= 8.1, p =.004 for the *sequence* storyboards).

The results of the recognition coding for the "What did you see happening in this scenario?" open response question indicate that, for the *sequence norm* storyboards: (1) more participant responses referenced the student's presentation of the proof in storyboards in which a norm was breached than in storyboards in which the norm was not breached; and (2) of those responses that contained a reference to the student's presentation of the proof, more responses contained evaluative references than non-evaluative references in the storyboards in which the norm was breached. For both the *details* and *sequence* norm storyboards, there were more negative evaluative references in storyboards that represented breaches of hypothesized norms and more positive evaluative references in storyboards that represented compliance with hypothesized norms.

One possible reason that the results of the *recognition* coding for all but storyboard 26004 of the *details norm* were not consistent with what would be expected from the transparency of the routine is the explicit way in which the teacher is represented as complying with the hypothesized norm in the control versions of the *details* norm storyboards. By contrast, the students' transcriptions of the proofs in the compliance versions of the *sequence* norm storyboards—like the teacher's use of the point of intersection between the angle bisectors of a parallelogram in the control version of storyboard 26004—is an action that fades into the background for those participants that recognize the student transcription practices depicted in the episodes as routine. If the student transcription of the proof were out of the ordinary, it would be reasonable to expect teachers to comment on this aspect of the scenario in the control versions of the *sequence* norm storyboards.

There is evidence in the recognition coding of the *sequence norm* open response data that suggests that, for some participants, the hypothesized-to-be routine transcription of a proof by a student was a departure from what was expected. Examples of such responses are the *negative evaluative references* that were present in control versions of the sequence norm storyboards. There were 12 such responses. One that illustrates that the student transcription of the proof was a breach of what was expected is: “Why didn’t the teacher have the student explain each step and talk about why it was correct or not?” (27005BC261-Control). The response goes on to say: “Explaining each step as they go along might be a better way to check to see if all students were understanding the concept”. This response directly mentions the manner in which the student transcribes the proof and, furthermore, states that it would have been “better” had the teacher required the student to explain each step as the student wrote the proof on the board. Other responses indicated that, rather than just write a proof on the board, the students should have been asked (or given an opportunity) to explain their proof.

On the one hand, the presence of responses like these indicate that the storyboards that represented compliance with the sequence norm were viewed as a departure from what was expected for at least some of the teachers that viewed the sequence norm storyboards. This could be seen as disconfirming evidence of the existence of the norm. On the other hand, there were comparatively few *negative evaluative references* to the control versions of sequence norm storyboards. What this suggests is that, for the majority (68/90) of responses to the control versions of the sequence norm storyboards, the student presentation of the proof was either (1) completely transparent (i.e., no reference to the presentation in the open response, 29/90) or else (2) referenced in a descriptive (rather than evaluative) way (i.e., non-evaluative reference to the student presentation, 39/90).

Both *positive* and *negative evaluative references* to the breach versions of the *details* and *sequence* norm storyboards provide evidence that participants recognized the breach of the norm.

The *positive evaluative references* of the breach provide evidence that the teacher's depicted actions in the storyboards that breach the norm are reasonable and warranted. In Chapter 4, I described the principle of *reasonable departure* as one of the guidelines for scripting storyboards in which the teacher's breach of the norm is a reasonable alternative to the expected normative way that a teacher would act in a situation. An example of a response that contains a *positive evaluative reference* to a *details* storyboard is: "A student was having trouble with a proof so the teacher told them what they were missing" (26003BC101-Breach). This was a response to a *more details* breach storyboard that indicates that the additional details the teacher requested from the student were "missing" from the proof otherwise. The positive evaluative reference suggests that the teacher's request to argue for the collinearity of the points was a reasonable action.

This response contrasts with a *negative evaluative reference* to the details provided by the teacher in the same storyboard: "The proof is getting lost in the minutia. If the teacher is concerned about the collinearity of the points, then include it in the given" (26003BC161-Breach). The negative evaluative response directly mentions that the teacher is asking for details that are more fine-grained than those that are expected in proofs. A *negative evaluative reference* to a *less details* storyboard that indicates the teacher breached what was expected by overlooking details that are hypothesized to be required is: "I don't like that the teacher said you didn't have to include the segments being congruent in the proof. You would have to include that to prove two triangles are congruent. A proof shouldn't leave any piece up to the imagination or interpretation." (26001BC291-Breach).

The *positive* and *negative evaluative references* to the students' presentations in the responses to the *sequence norm* storyboards also provide evidence that participants recognized the breach of the norm. The *positive evaluative references* to the breach indicate that the teacher's actions in the breach versions of the storyboards were reasonable departures from the hypothesized norm. One participant wrote: "The instructor might have been attempting to determine if the student really understood the proof by changing the labeling" (27001BC6-Breach). This was a response to a breach version of a storyboard that targets the *labels* aspect of the *sequence* norm. In this storyboard, the student completes the transcription of the proof but forgets to write the labels on the diagram. Rather than allow the student to add the labels after the fact—as is depicted in the control version of the storyboard—the teacher labels the diagram using different labels than the ones that the student wrote in the proof.

Relabeling the diagram was a rather strong move on the part of the teacher toward breaching the norm that students are expected to transcribe, rather than recreate, proofs when writing them on the board. Other responses reacted negatively to this move, for example: "The teacher used a pretty extreme way of illustrating why it is important to label" (27001BC221-Breach). Another participant wrote: "A stressed-out student because he had to re-do his proof since the teacher put different labels on it" (27001BC21-Breach). These example *negative references* indicate that participants reacted to the breach of the norm. But even this move on the part of the teacher to breach the norm—by challenging the validity of the student's transcription of the proof—was not seen as something that was totally unreasonable, as evidenced by the *positive evaluative reference* cited above.

There were comparable responses to storyboards that depicted the teacher breaching the *sequence* norm by taking issue with the order in which the *markings* and *reasons* of a proof were copied by a student. An example of a *positive evaluative reference* to a *sequence (markings)* storyboard that breaches the norm is: "The teacher did well to hold the student accountable to every step in the proof. It is fair to say that you should not mark the diagram until you have proven that marking" (27003BC271-Breach). An example of a *positive evaluative reference* to a breach of the *reasons* aspect of the *sequence norm* is: "A student was presenting the mathematics incorrectly and the teacher immediately corrected him before he formed a bad habit" (27006BC241). Such responses that included *positive evaluative references* to the teacher's interference with the student transcription of the proof were by no means common or typical of the responses that participants provided. But they are valuable nonetheless, because they provide evidence that the responses to the questions that followed the breach versions of the storyboards were not the result of those storyboards representing teaching that was unreasonable or unjustifiable. Rather, these responses—consistent with the ethnomethodological design of the study—represent participants reactions to the unexpected teaching that was depicted in the breach versions of the storyboards.

Findings of the recognition coding for the first question

The results reported above indicate that, in general, participants recognized the teaching that was depicted in the control versions of the storyboards as routine. This was especially evident in the responses to storyboard 26004 and the control versions of the *sequence norm* storyboards. In these storyboards, the expected-to-be-routine actions of the teacher with respect to the necessary details of the proof and the student presentation of a proof to the class were generally not mentioned by participants in their open responses to the "What did you see happening in this scenario?" question. That the majority of the open responses contained no reference to those aspects of the situation that were hypothesized to be normative is consistent with the ethnomethodological principle of the transparency of the routine.

Even in the cases of the *details* storyboards in which the teacher's handling of the details of the proof were explicitly mentioned by participants, it was the case that, in the instances of the teacher breaching the hypothesized normative ways of handling the details of the proof, the teacher's handling of the details of the proof was viewed negatively by participants. By contrast, the teacher's handling of the *details* of the proof in the storyboards that comply with the hypothesized norm was viewed positively by participants. The results of the *recognition* coding for both the *details* and *sequence* norm storyboards indicate that participants noticed the breaches of the hypothesized norms.

Results of the *recognition coding* for the other open response questions

There were three other open response questions that participants answered after viewing each storyboard. These were follow up questions to the *episode appropriateness* and the two *segment of interest appropriateness* rating questions. After rating the appropriateness of the teacher's review of the proof, the appropriateness of the teacher's actions during a segment of the storyboard in which the norm is (or is not) breached, and the appropriateness of the teacher's actions in some other segment of the storyboard, participants were provided with an open response field. The prompt for this open response field was: "please explain your rating".

The responses to these fields were also coded for *recognition.* I report the results of the analysis of the responses to these questions separately because these questions prime the participant to evaluate what is happening in the episode. The responses to these questions also provide evidence about participants different responses to the breach and control versions of the storyboards.

Results of recognition coding for the *episode appropriateness* question

The *episode appropriateness* rating question is the most general of the three follow up questions. It is reasonable to hypothesize that participant responses to this follow up question would be generally consistent with the responses reported for the "What did you see happening in this scenario?" question. That is, responses to storyboards which represent a breach of a hypothesized norm will have more *references* to the target norm than responses to storyboards in which the norm is not breached. Since participants are being asked to rate the appropriateness of the work of the teacher, it is reasonable to suppose that responses that contain *references* to either the details of the proof or the student's presentation of the proof will contain more *evaluative references* than *non-evaluative references*. This is likely to be the case regardless of condition, because participants are being asked to explain their rating of the teacher's review of the proof in the scenario. Finally, it is reasonable to suppose that there will be more *positive evaluative references* in storyboards that represent the teacher complying with a hypothesized norm than in storyboards in which the teacher breaches a hypothesized norm. Similarly, it is reasonable to suppose that there will be more *negative evaluative references* in storyboards that represent the teacher breaching a hypothesized norm. These hypotheses are investigated during the reporting of the results, below.

The results of the recognition coding for the follow up to the *episode appropriateness* rating question are reported in Table 38-Table 40. These tables are structured the same way as those used to report the results of the *recognition coding* for the "What did you see happening in this scenario?" question. I discuss the results of the *recognition coding* for the follow up to the *episode appropriateness* rating question after presenting the tables.

Table 38: Counts of responses that contained a reference to the target norm for the second open response question, by norm and condition.

Norm/Condition	No Reference	Reference	Total
Details/Breach Storyboards	26	32	58
Details/Control Storyboards	37	19	56
Sequence/Breach Storyboards	30	58	88
Sequence/Control Storyboards	57	33	90

The results reported in Table 38 are consistent with those reported for the first open response question. A chi-square test of association for the *references* and *no references* counts for the *details* and *sequence* norms indicates a significant association between the number of responses with *references* and the breach/control condition (χ^2= 5.2, p =.02 for the *details* storyboards; χ^2= 15.32, p <.001 for the *sequence* storyboards). As predicted, these results indicate that the responses to the storyboards that breach the norm have more responses that *reference* the details of the proof or the student presentation of the proof.

Table 39: Counts of responses that contained evaluative references to the target norm for the second open response question, by norm and condition.

Norm/Condition	Evaluative	Non-evaluative	Total
Details/Breach Storyboards	28	4	32
Details/Control Storyboards	19	0	19
Sequence/Breach Storyboards	52	6	58
Sequence/Control Storyboards	24	9	33

The results reported in Table 39 are consistent with the results of the analysis for the first open response question. A Fisher's exact test of association indicates that for the *details* storyboards, the number of responses that contained *evaluative references* was independent of the condition (breach/control) of the storyboard (*p* =.283 for the *details* storyboards). For the *sequence* storyboards, a chi-square test of association indicates a significant association between the condition (breach/condition) of the storyboard and the presence of *evaluative* or *non-evaluative references* (χ^2= 4.38, *p* <.036).

For both sets of storyboards, there were more *evaluative references* than *non-evaluative* references. These results are in line with the hypothesis presented above. Because the response field prompted participants to rate the appropriateness of the teacher's actions, it was reasonable to suppose that there would be more evaluative references than non-evaluative references in both the breach and control conditions. The results reported in Table 40 bear this out.

Table 40: Counts of responses that contained positive and negative evaluative references to the target norm for the second open response question, by norm and condition.

Norm/Condition	Positive Reference	Negative Reference	Total
Details/Breach Storyboards	2	26	28
Details/Control Storyboards	16	3	19
Sequence/Breach Storyboards	4	48	52
Sequence/Control Storyboards	5	19	24

For the *details* norm storyboards, the results reported in Table 40 were in line with the prediction that there would be more *negative evaluative references* in storyboards that represented a breach of a hypothesized norm. A chi-square test shows that, for the *details* storyboards, there was a significant association between the breach/control condition of a storyboard and the polarity of an *evaluative reference* ($\chi^2 = 28.45$, $p < .001$). In the case of the *sequence* storyboards, there were more *negative evaluative references* in storyboards that depicted a breach of the hypothesized norm. However, a Fisher's exact test indicates that there is no significant association between the polarity of an *evaluative reference* and storyboard condition for the *sequence* storyboards ($p = .13$). It was also the case that there were more negative *evaluative references* than *positive evaluative references* in the control versions of the *sequence* storyboards. This is counter to the prediction that there would be more *positive evaluative references* than *negative evaluative references* in control storyboards.

An example of a *negative evaluative reference* to a control version of a *sequence* storyboard is: "If a teacher has a student put a problem on the board, the student should be explaining what they are doing as they go along" (27006 BC22-Control). This reference to the student presentation of the proof was coded as *negative evaluative reference* because it states that what the student actually did—i.e., transcribe the proof onto the board without addressing the class—is counter to what the teacher should have had the student do. A different example of a *negative evaluative reference* to the control version of a *sequence* storyboard is: "Did not let Eta explain the proof" (27001BC292-Control).

Each of these examples indicates that more was required of the student than a simple transcription of the proof. In the first case, the response suggests that a student should coordinate speaking about the proof while writing the steps of the proof on the board. This response clearly takes issue with the fact that the student transcribes the proof, because it should be explained as it is developed. For such an explanation to be sensible, it would be necessary for the proof to be written in a logically coherent way on the board. In the second case, the response indicates that the proof needs to be explained by the student. This response stops short of indicting the transcription itself. It could be that the participant would have asked the student to explain the steps in the proof after the proof had been transcribed (as was the case for the students that explained the proofs after transcribing them, described in Chapter 3).

The example responses considered above indicate that, for some participants, the expected-to-be-routine act of a student transcribing a proof was, itself, viewed as breach. I argued above that the existence of a small number of responses like these in the corpus strengthens the results of study. That some participants explicitly marked the student's transcription of the proof as a departure from what they would expect students to do, suggests that, for those participants that did not mark the student presentation in this way, the student's presentation was routine.

Results of recognition coding for the *segment appropriateness* questions

In addition to the analysis of the general open response prompts reported above, participants' responses to the *segment appropriateness* rating questions were also coded using the *recognition* scheme. The *segment of interest appropriateness* rating question zoomed in on the segment of the storyboard in which the teacher breaches or does not breach the norm. The *distracter segment appropriateness* rating question zoomed in on some other segment of the storyboard that was not related to the target norm. Because the *segment of interest* rating question focused specifically on the moment in a storyboard where a teacher acts in accordance with or else departs from a hypothesized norm, it is reasonable to suppose that responses to the follow ups to these rating questions would be consistent with those reported for the *episode appropriateness* rating questions. One difference between these two rating questions, though, is the dependence of the *segment of interest* rating questions on the segment of the storyboard that was attached to the question. Since the action that participants are rating is immediately present and represented already in the selected frames of the storyboard, it is possible that there are fewer responses—when compared to the totals for the *episode appropriateness* rating question—that *reference* the action that takes place in the segment that is being rated.

I report first the results of the *segment of interest* appropriateness rating question. The reporting of the results for the responses to this question follows the same structure as the report of the results for the first two open response questions. I then report the results of the *distracter segment* appropriateness rating question. Since this question showed participants a segment of the storyboard that was generally unrelated to the target norm, it is reasonable to expect that responses to these questions which contain few references to either the details or sequence norm. The results for the *distracter segment* rating question indicate this to be the case.

Table 41 reports the results of the *references* coding for the follow up to the *segment of interest* appropriateness rating question.

Table 41: Counts of responses that contained a reference to the target norm for the third open response question, by norm and condition.

Norm/Condition	No Reference	Reference	Total
Details/Breach Storyboards	21	37	58
Details/Control Storyboards	25	31	56
Sequence/Breach Storyboards	31	57	88
Sequence/Control Storyboards	54	36	90

There were roughly equal numbers of responses that contained a *reference* to the *details* norm in the follow up to the segment of interest appropriateness rating. A chi-square test of association indicates that there is no significant relationship between storyboard condition and whether a response contained a *reference* for the *details* storyboards. Responses to the follow up to the *segment of interest* appropriateness question for the *sequence* storyboards contained more references to the student presentation of the proof in storyboards in which the teacher was depicted as interfering with the student's transcription of the proof. A chi-square test of association indicates there is a significant relationship between storyboard condition and whether a response contained a *reference* to the student presentation of the proof (χ^2=10.94, p <.001)

Table 42: Counts of responses that contained evaluative references to the target norm for the third open response question, by norm and condition.

Norm/Condition	**Evaluative Reference**	**Non-evaluative Reference**	**Total**
Details/Breach Storyboards	33	4	37
Details/Control Storyboards	26	5	31
Sequence/Breach Storyboards	55	2	57
Sequence/Control Storyboards	26	10	36

Table 42 reports the breakdown of *evaluative* and *non-evaluative references* for the *details* and *sequence* storyboards, by condition. These tallies are consistent with the general pattern of responses for the two more general open response questions that were reported above. Table 43 indicates that there were more *negative evaluative references* in storyboards that breach a hypothesized norm, and there were more *positive evaluative references* in storyboards in storyboards in which no norms are breached.

Table 43: Counts of responses that contained positive and negative evaluative references to the target norm for the third open response question, by norm and condition.

Norm/Condition	**Positive Reference**	**Negative Reference**	**Total**
Details/Breach Storyboards	3	30	33
Details/Control Storyboards	23	3	26
Sequence/Breach Storyboards	5	50	55
Sequence/Control Storyboards	12	14	26

In the case of the *details* storyboards, the association between breach/control condition and the polarity of the evaluative reference is significant ($\chi^2 = 37.7$, $p < .001$). The association between condition and polarity of the *evaluative references* was also significant for the *sequence* storyboards ($\chi^2 = 14.62$, $p < .001$). The results from the analysis of the *segment of interest* follow up question are consistent with those reported from the general open response and *episode appropriateness* rating questions reported above.

The *distracter segment* rating questions asked participants to rate the actions of the teacher in a segment of a storyboard that was generally unrelated to the target norm for the storyboard. It is therefore reasonable to hypothesize that there would be few references overall to either the *details* or *sequence* norm, and that this would be the case across both conditions. Table 44 reports the results of the recognition coding for the follow up to the *distracter segment* rating question.

Table 44: Counts of responses that contained a reference to the target norm for the fourth open response question, by norm and condition.

Norm/Condition	No Reference	Reference	Total
Details/Breach Storyboards	53	5	58
Details/Control Storyboards	49	7	56
Sequence/Breach Storyboards	73	15	88
Sequence/Control Storyboards	79	11	90

For both the *details* and *sequence* storyboards, there were few references to the target norm in responses to the *distracter segment* rating question. In both cases, a chi-square test of association indicates that there was no significant relationship between storyboard condition and whether a response contained a reference to the target norm. This result is consistent with what was hypothesized for the *distracter segment* rating questions.

Results of the *recognition* coding for the three follow up questions

The results reported above for the *recognition* coding for the open responses to the *episode appropriateness* and *segment of interest appropriateness* rating questions were consistent with the results reported for the analysis of the general open response question. These results indicate that breaches of hypothesized norms in the storyboards were noticed by participants. The open responses to the *distracter segment appropriateness* rating question contained few responses that referenced the target norm and these responses were equally likely across breach/control storyboard condition.

The purpose of the *recognition* coding was to analyze responses for evidence that the response referenced the details of a proof or a student presentation of a proof. The analysis of the open response data for these specific references to what is at stake in each hypothesized norm provide evidence that participants reacted to the breaches of the hypothesized norms. The next section reports the results of the more general appraisal coding. The more general appraisal coding complements the more specific *recognition* coding by analyzing response for evidence of positive or negative evaluations about any aspect of the storyboards.

Results of appraisal coding

All of the open responses in the corpus were coded using the appraisal scheme that was described above. The purpose of coding the responses for evidence of appraisal was to capture the more general reactions that participants had to the different storyboards. Consistent with the design of a breaching experiment, I hypothesized that, in the case of storyboards that represent a breach of a norm, participants would tend to react negatively to a storyboard. By contrast, the control versions of the storyboards were designed to represent routine teaching. A reasonable hypothesis based on the assumption that the control storyboards represent routine teaching is that they would contain roughly equivalent numbers of positive and negative evaluations.

Each response in the corpus was coded for judgments, appreciations, or statements of affect, and each instance of a attitudinal appraisal was coded in each responses. This means it was possible for a response to contain multiple instances of the same kind of statement of attitude (e.g., there could have been several judgments of a teacher), as well as instances of different kinds of statements of attitude (e.g., a judgment of the students, an appreciation of the proof), and statements of attitude that had different polarities (e.g., a response could contain both positive and negative judgments of the teacher).

Table 45 shows the total number of positive[55] and negative statements of attitude throughout the open responses in the corpus. The results are reported according to storyboard condition. The entire corpus contained 1168 open responses to the 4 different open response questions for each of the 20 storyboards. Throughout the entire corpus, there were equal numbers of responses to the breach and control conditions. In each case, there were 584 total responses. These were divided equally (146 per question) among the 4 open response questions that were described above. These were the general open response question and the responses to the follow up prompts that accompanied the *episode appropriateness,* the *segment of interest appropriateness,* and the *distracter segment* appropriateness rating question. The results reported in Table 45 are for the entire corpus across all 4 question types. Furthermore, the results reported in Table 45 are simple counts of the number of statements of positive or negative attitude that were coded in the open responses to the storyboards in the breach and control conditions.

Table 45: Counts of statements of attitude, tallied by polarity and storyboard condition.

Storyboard Condition	Statements of Positive Attitude	Statements of Negative Attitude	Total
All Breach Storyboards	211	473	684
All Control Storyboards	309	310	619

The results reported in Table 45 are consistent with the hypotheses stated above. The open responses to the storyboards in which a hypothesized norm was breached had more statements of negative attitude than statements of positive attitude. In the case of the control storyboards, there were nearly equal numbers of positive and negative statements of appraisal. A

[55] Throughout the corpus of open responses (n = 1168), there were 18 responses that contained mathematical appreciations that were coded as *neutral*. These statements were tallied as *positive* appraisals for the results reported in this section.

chi-square test indicates that there is a significant association between storyboard condition and the number of positive or negative statements of attitude ($\chi^2 = 48.49$, $p < .001$).

The results reported in Table 45 provide evidence that the control versions of the storyboards did, in fact, capture routine teaching. This follows from the nearly equal numbers of positive and negative evaluative statements in the open responses to the control versions of the storyboards. Furthermore, the comparatively large proportion of the attitude statements to the breach versions of the storyboards that were negative statements of attitude suggest that the breach versions of the storyboard presented participants with departures from what was expected.

The unit of analysis for the results reported in Table 45 is a statement of attitude. This means that each statement of attitude in a response was included in the totals. To further investigate the relationship between attitude polarity and storyboard condition, I recoded the data to eliminate multiples, by polarity, within each response. Thus, if a response contained 3 positive statements of attitude and 2 negative statements of attitude, it was recoded as (1) for *positive attitude* and (1) for *negative attitude*. Table 46 reports tallies of statements of attitude after applying this reduction. The unit of analysis for the results reported in Table 46 is an open response (n = 584 for each storyboard condition).

Table 46: Counts of open responses that contain positive or negative statements of attitude, by storyboard condition.

Storyboard Condition	**Statements of Positive Attitude**	**Statements of Negative Attitude**	**Total**
All Breach Storyboards	176	352	528
All Control Storyboards	248	245	493

The results reported in Table 46 are consistent with those reported above. Across the corpus for the control storyboards, there were 248 responses that contained at least one statement of positive attitude and 245 responses that contained at least one statement of negative attitude. By contrast, across the corpus for the breach storyboards, there were 176 responses that contained at least one statement of positive attitude compared to 352 responses that contained at least one statement of negative attitude. A chi-square test of association indicates that there is a significant relationship between storyboard condition and attitude polarity ($\chi^2 = 29.54$, $p < .001$).

The results reported in Table 46 are refinement of the results reported in Table 45 because multiples have been eliminated. The results reported in Table 47 (below) refine these results further by distinguishing 4 categories of response: those that contain only positive statements of attitude, those that contain only negative statements of attitude, those that contain both positive and negative statements of attitude, and those that contain no statements of attitude.

Table 47: Counts of open responses that contain only positive attitude, only negative attitude, both, or neither.

Storyboard Condition	Statements of Positive Attitude	Statements of Negative Attitude	Both Positive and Negative Attitude	None	Total
All Breach Storyboards	78	254	98	154	584
All Control Storyboards	166	163	82	173	584

The results reported in Table 47 are consistent with those reported in Table 45 and Table 46. The breach storyboards contained more responses that contained only negative statements of attitude than responses that contained only positive statements of attitude. By contrast, the control storyboards contained nearly equal numbers of responses that contained only positive statements of attitude and responses that contained only negative statements of attitude. A chi-square test of association indicates a significant relationship between storyboard condition and the categories of attitude in Table 47 ($\chi^2 = 54.12$, $p < .001$)

The results reported above provide evidence to support the hypotheses stated at the beginning of this section. Throughout the corpus, the responses to the breach versions of the storyboards yielded more negative statements of attitude than positive statements of attitude. The fact that responses to the breach versions of the storyboards produced more negative statements of attitude is consistent with the results of the breaching experiments conducted by Garfinkel (1963) and the virtual breaching experiments conducted by others (Herbst, Aaron, Dimmel, & Erickson, 2013). The higher number of negative statements in the open responses to the breach versions of the storyboards suggest that participants were reacting to the breaches of the norm, as opposed to some other aspect of the storyboards.

The claim that participants' negative reactions to the treatment storyboards are a result of the designed breaches of the hypothesized norms, as opposed to some other aspects of the storyboards, is strengthened by the matched design of breach/control storyboard pairs and the results of the appraisal coding for the control storyboards. In contrast with the breach storyboards, responses to the control storyboards contained roughly equal numbers of positive and negative statements of attitude. Because the breach and control storyboards were the same except during those frames where the teacher is shown breaching (or complying with) a hypothesized norm, it follows that the teacher's breach of the norm is what prompted participants to react negatively to the storyboard.

In Chapter 5, the closed-responses to the *distracter segment* rating question were used as a within storyboard control. The *distracter segment* rating question focused on the *common* *ground* segment of each breach/control storyboard pair. The *common ground* was a set of frames in the storyboard that were common across breach and control versions. Examining the polarity of the attitude statements that participants expressed in responses to the *distracter segment* open response field provides additional supporting evidence that participants noticed and reacted to the breaches of the norm in the breach versions of the storyboards.

The action that takes place in the *common ground* of the breach/control storyboards was generally not related to the target norm that was breached (or not breached) in the storyboards. Furthermore, the results presented thus far indicate that participants' negative reactions to storyboards are linked to breaches of the hypothesized norms. For these reasons, it is reasonable to hypothesize that, for the open responses to the *distracter segment* rating questions, there will be roughly equal numbers of responses that contain positive and negative statements of attitude. Table 48 reports the results of the reduced attitude coding for responses to the *distracter segment* rating question.

Table 48: Polarity of statements of attitude contained in open responses to the the *distracter segment* rating question, by condition.

Storyboard Condition	Statements of Positive Attitude	Statements of Negative Attitude	Total
All Breach Storyboards	53	64	117
All Control Storyboards	62	61	123

The results reported in Table 48 are consistent with the hypothesis stated above. There were slightly more responses to the *distracter segment* open prompt that contained statements of negative attitude than statements of attitude. There were also roughly equal total numbers of responses that contained statements of attitude to the *distracter segment* rating question. There is no significant relationship between storyboard condition and the polarity of an attitude statements. These results further corroborate the claim that participants were reacting to the planned departures of hypothesized norms when they reacted negatively to the breach versions of the storyboards.

Summary of the analysis of the open responses

In this Chapter I defined two schemes for coding the open response data that was gathered with the instrument described in Chapter 4. The open responses were coded for evidence that the response recognized the target semiotic norm, using a network of dichotomous codes for *references* to the target semiotic norms. The open responses were also coded for evidence that participants made evaluative statements in their open responses. The open response data was indexed using a system that blinded the responses to breach/control condition. The indexed data was coded using the *recognition* and *attitude appraisal* schemes.

Results of the coding for responses to storyboards in which a norm was breached were compared to the results of the coding for responses to storyboards in which no norms were breached. For both the *recognition* coding and the coding for positive and negative statements of attitude, there were significant differences across breach and control conditions. These results were generally consistent with the results of the analysis of the closed-response data that was reported in Chapter 5. In the case of the *recognition* coding, participants made more references to the target norms in storyboards in which a norm was breached when compared to storyboards in which no norms were breached. This pattern of responses indicates that the teaching represented in the breach versions of the storyboards represented departures from what was expected. In the case of the attitude coding, the storyboards in which a norm was breached yielded higher numbers of responses that contained negative statements of attitude. This result contrasted with the attitude coding for the control storyboards. The control storyboards yielded approximately equal numbers of positive and negative statements of attitude.

The analysis of the open response data provides evidence that the *details* and *sequence* norms are recognized by secondary teachers as routine aspects of the work of doing proofs in geometry. In the next chapter, I review the work that was conducted to hypothesize and test for secondary teachers' recognition of semiotic norms. I discuss the findings from the different parts of the study and consider the scholarly and pedagogical significance of this research.

CHAPTER 7:

DISCUSSION OF FINDINGS AND CONCLUSION

The research conducted for this dissertation was an investigation of the routines that secondary teachers use to regulate student work when doing proofs in geometry. The study consisted of two parts. The first part of the study analyzed video records of doing proofs in geometry classrooms to ground hypotheses of what I call *semiotic norms.* The second part of the study used representations of geometry instruction to investigate the extent to which secondary teachers recognize the semiotic norms that I call *details* and *sequence*. Below, I review each part of the study and highlight the findings. Following this review, I discuss the significance and implications of this research before offering concluding remarks about the study.

Review of the description of episodes of doing proofs

The study reported in Chapter 3 was a multimodal description of how proofs are presented by students and checked by teachers in geometry classrooms. To investigate the routine ways that students present proofs to geometry classes, I analyzed video records of student public presentations of proofs to their peers. To investigate the routine ways that teachers check the proofs that students produce in geometry classrooms, I analyzed video records of teachers scrutinizing the details of proofs.

In the case of presenting and checking proofs in geometry classrooms, I hypothesized that there are normative ways that secondary teachers expect students to use semiotic resources. The analysis of video episodes of student presentations of proofs suggested that, for students, the activity of presenting a proof to a class consisted primarily of reproducing an already completed proof onto a central classroom board. This is not to suggest that no students engage in *chalk talk*, or that there are no opportunities for students to present proofs to the class without first transcribing them; rather, the video records of student presentations of proofs provided initial evidence that students are generally not expected to coordinate speaking, writing, drawing, and gesturing when presenting proofs the class. Students were generally not expected to engage in *chalk talk*—the disciplinary practice of generating a proof through the coordinated use of speaking, gesturing, writing, and drawing (Artemeva & Fox, 2011)—but rather engaged in an activity that I described as *proof transcription.* Unlike *chalk talk,* during proof transcription, the proof that is being transcribed is not required to be a coherent mathematical entity until the transcription is completed.

In some cases, reproducing the proof on the board was sufficient for students to complete the work that was expected by the teacher. In other cases, students were called to the board to verbally explain the written work that had been previously transcribed onto the board. Finally, I examined a case of a student presenting a proof to the class that the student had not previously completed. A commonality across these different instances of students presenting proofs to a geometry class is the isolated or redundant use of semiotic resources by students doing the presenting.

I use the words *isolated* and *redundant* in specific ways. By “isolated” I mean using semiotic resources from one communicative mode at a time. By “redundant” I mean using resources from different modes to do the same communicative work. An example of the isolated use of semiotic resources is the tendency for students to silently transcribe a proof onto the board, or to speak about the steps of a proof only after the transcription has been completed.

Even in the case of the student that generated a proof for the class—in what could be the closest approximation to the disciplinary practice of *chalk talk*—there was almost no simultaneous use of semiotic resources by the student. The student stated a step verbally. Next, the student wrote the step after it had been stated. Next, the student marked the step in the diagram. The student used one modality at a time to present the proof to the class.

An example of the redundant use of semiotic resources is the use of deictic gestures co-expressed with diagrammatic speech to verbally read a statement that has been written in a proof. The redundant use of semiotic resources was observed in students that were called to the board to explain their proofs after the transcription had been completed. A student might, for example, verbally read the written statement: "$\overline{AB} \cong \overline{CD}$" by saying: "segment AB is congruent to segment CD" while pointing to "segment AB" and then "segment CD" in a diagram. In this example, the use of labels on the diagram and in the written statement of the proof overlaps with the pointing gestures, and the verbalization of the step overlaps with what is written.

The student tendency to use semiotic resources redundantly or in isolation contrasts with how mathematical experts use semiotic resources. For instance, a teacher might verbalize the same step written above by using conceptual speech (e.g., "the opposite sides are congruent") that is co-expressed with non-pointing gestures (e.g., a metaphorical gesture that indicates an equivalence of opposite things). In such an ensemble, the labels in the diagram and in the written statement of the proof achieve the linking of the statement to the diagrammatic object to which it refers, which frees up the teacher to use speaking and gesturing to do different communicative work. This additional work is achieved by the use of conceptual speech and metaphorical gestures. The conceptual verbalization of the step links the specific realization of the objects in the diagram to more general mathematical concepts, and the co-expression of the conceptual speech with a metaphorical gesture provides an additional representation of the congruence relation between the segments. The coordinated, synergistic use of multimodal resources is a disciplinary communication practice that is routinely displayed by mathematical experts (Artemeva & Fox, 2011; Greiffenhagen, 2014).

Although the setting of presenting a proof to a geometry classroom creates an opportunity for students to develop the disciplinary communication practices mathematical experts display during *chalk talk,* the episodes in the video corpus indicated that students were not expected to use such practices when presenting proofs to geometry classrooms. Since students are novices, rather than mathematical experts, it is not surprising that they did not spontaneously offer presentations of proofs that were on par with the practiced routines of *chalk talk* that are used by mathematicians. That there are disciplinary specific communication skills that students need to learn is not the finding; the finding is that teachers did not seem to recognize that the task of presenting a proof to a geometry classroom presents an opportunity for students to hone discipline-specific communication practices.

From my observations of the video episodes of students presenting proofs to the class, I hypothesized that the practice of *proof transcription* was part of the routine work that secondary teachers expect students to do when they are asked to present a proof to the class. I also observed that during the transcription of a proof, the order in which the different parts of the proof were transcribed was immaterial to the completion of the work. Students might complete a transcription without labeling the diagram that accompanies the proof (rendering the written arguments unintelligible), or write the statements and reasons before drawing the diagram or writing what was given or the statements to be proved.

Based on these observations, I hypothesized that the *sequence* in which the different parts of a proof are transcribed when a student presents a proof in geometry need not be logically coherent. A transcription of the different parts of a proof could proceed in whatever order the student found convenient. I used storyboards to create representations of classroom scenarios that depict breaches of the *sequence norm* during the experimental part of the study. The development of these storyboards and the findings from the experimental study are summarized below.

In the case of video episodes of teachers checking proofs, I observed that there were certain kinds of details that teachers expected to be included in the written statements and reasons of a proof. Some of the details that are necessary involve making distinctions between geometric objects and geometric properties. An example of one such needed detail is drawing the distinction between an angle (object) and its measure (property). Other details that are needed involve methodically unpacking the written information that is given into statements that are warranted by definitions or theorems. For example, if one of the givens for a proof problem is "C is the midpoint segment AB", the congruence of segments AC and CB needs to be explicitly established in the written statements and reasons of the proof before it could be used to warrant other claims. Teachers would expect students to write a statement that identifies that C is the midpoint of AB (warranted by what is given) and a separate statement that AC and CB are congruent (by definition of midpoint).

In the video episodes of teachers scrutinizing the details of proofs, teachers required students to unpack written statements (e.g., $\angle 1 \cong \angle 4$, C is the midpoint of $\overline{AB}$) into other written statements (e.g., $m\angle 1 = m\angle 4$, $\overline{AC} \cong \overline{CB}$) when using those statements in a proof. Teachers generally did not require such unpacking for statements of co-exact (existence of points of intersections, the separation relations among regions) and other properties (such as incidence or collinearity) that were warranted by the diagram that accompanied a proof. Students were not, for example, asked to unpack the visually stated betweenness relations that held among the points in a diagram into written statements in the proof that would be warranted by primitive postulates or definitions. The properties that were given tacitly in the diagram were not required to be explicitly stated as details of the proof, even though teachers expected students to unpack the properties that were given explicitly through written statements into more basic steps.

The regularity of these expectations across instances of teachers checking proofs in geometry classrooms lead me to hypothesize that teachers scrutinize the details of proofs in mathematically specific ways that are linked to different semiotic resources (e.g., writing, the diagram). I referred to these routines as the *details norm.* The experimental part of the study used representations of geometry instruction to investigate the extent to which secondary teachers recognize the routine ways that teachers check the details of proofs.

The multimodal descriptions of proofs being presented and proofs being checked allowed me to observe regularities of these activities as they played out in different geometry classrooms. The regularities that I observed concerned how teachers expect students to manage semiotic resources when dong proofs in geometry. I used the term *semiotic norm* to describe these regularities. The hypothesized routine ways that teachers expected students to use semiotic resources when doing proofs in geometry were based on my observations of video episodes of geometry classrooms. The second part of the study attempted to determine whether these regularities are recognizable to secondary mathematics teachers in general.

Review of the virtual breaching experiment

The research design described in Chapter 4 was a *virtual breaching experiment with control.* The purpose of the instrument was to investigate whether the *semiotic norms* that were hypothesized in Chapter 3 could be empirically studied. The instrument operationalized the ethnomethodological principle of the transparency of the routine. The underlying hypothesis of the empirical study was that teachers would react to representations of geometry classrooms that depicted departures from hypothesized semiotic norms.

The notion that what is routine is generally transparent was made operational for the study by using matched pairs of storyboards as probes. The storyboards were shown to different groups of secondary teachers. The 73 secondary mathematics teachers that participated in the study were randomly assigned to treatment and control conditions. In each condition, participants viewed a set of storyboards that represented geometry classrooms. The storyboards were the probes.

The storyboards were created using a customizable graphic language that had been previously validated as an effective medium for representing mathematics classrooms. Storyboards were created in matched pairs. Each matched pair of storyboards were minimally different from each other. A minimally different matched pair featured the same teacher, saying mostly the same things, to the same students, working on the same math problem. Where the storyboards were different was during a 3-5 frame segment, during which the teacher in a storyboard was represented as acting in ways that complied with (control) or departed from (breach) a semiotic norm that was hypothesized in Chapter 3. The semiotic norms that were investigated for the empirical study were the *details norm* and the *sequence norm.*

I created four matched pairs of storyboards to investigate the extent to which secondary teachers recognized the *details* norm. I created six pairs of matched storyboards to investigate the extent to which secondary teachers recognized the *sequence* norms. The different storyboards in a set realized different ways that a target norm could be breached by the teacher in a geometry classroom. The teacher was represented as breaching the *details norm* by (a) accepting a proof that did not include a detail that was hypothesized to be needed, and (b) insisting that a proof include a detail that was hypothesized to be unnecessary. The teacher was represented as breaching the *sequence norm* by (a) stopping a student's transcription of a proof and asking the student to explain the steps in the proof as they are being written on the board, and (b) requiring a student to re-do a proof because some parts of the proof were transcribed out of order.

Participants in the different conditions viewed either the breach version or the control versions of the storyboards that were in a breach/control matched pair. Each participant in the study viewed two breach storyboards and two (unrelated) control storyboards. After viewing the storyboards, participants were asked a series of open- and closed-ended questions. The purpose of asking participants both open and closed-ended questions was to provide complementary data sources to investigate teachers' reactions to the storyboards.

Responses to closed-ended rating questions—using a 6-point Likert-like response format for appropriateness—were analyzed using independent (across-condition) and paired-samples (within-condition) *t*-tests. Responses to the open ended questions were coded for evidence that participants referenced the target semiotic norm or made positive or negative evaluations of the storyboard. The analysis of the closed-ended questions indicated that, in general, participants rated lower the work of the teacher in those storyboards that depicted a breach of a semiotic norm. Concordantly, the analysis of the open-response questions indicated that, in general, participants significantly noticed the work of the teacher with respect to the semiotic norms in storyboards that represented breaches of the hypothesized norms and failed to notice the work of the teacher with respect to the semiotic norms in storyboards that represented compliance with the hypothesized norms.

Findings

The purpose of this dissertation was to investigate the following questions:

1. How do students use semiotic resources when doing proofs in geometry ?
2. To what extent do secondary teachers recognize routine ways semiotic resources are used to do proofs?

The multimodal analysis of the video episodes of teachers and students presenting and checking proofs in geometry classrooms reported in Chapter 3 answered the first research question. The findings of this part of the study were a set of *semiotic norms.*

The *channel norm* is the label I used to describe the fact that the written statements and reasons of a proof are the primary semiotic resources that teachers expect students to use when doing proofs in geometry. The *channel norm* manifests in student presentations of proofs to the class. When students present proofs to the class, the teacher holds the student accountable to produce a written record of the proof that is complete and correct. Students comply with this expectation by creating a mark-for-mark reproduction of a proof they have already completed, through an act that I described as *proof transcription.* The *sequence norm* is the label I used to describe the fact that students transcribe proofs in ways that facilitate their reproduction of the proof and not necessarily in ways that are mathematically sensible. Finally, when the teacher scrutinizes the details of a proof, the teacher checks to make sure that claims that can be warranted from written statements are explicitly unpacked into other written statements. Statements that are warranted by the diagram do not require such unpacking. The *details norm* is the label I used to describe the routine ways that the details of proofs are scrutinized by teachers in geometry classrooms.

The *channel, sequence,* and *details* norms hypothesized in Chapter 3 were observed regularities in a sample of geometry classrooms. The second research question concerned whether these local observations of classroom regularities would be recognizable by secondary teachers as normative aspects of the situation of doing proofs in geometry. The study that was described in Chapter 4 answered this research question. Throughout the study, storyboards in which the teacher in the classroom was depicted as acting in ways that departed from a hypothesized semiotic norm were rated lower and produced more negative evaluative statements than storyboards in which the teacher was depicted as complying with the hypothesized regularities. The pattern of lower appropriateness ratings and the greater number of negative evaluative comments for the breach storyboards in the study—combined with the matched breach/control design of the experiment—suggest that the hypothetical norms were recognized by secondary teachers.

The matched breach/control design of the study bolsters the claim that participating teachers were actually reacting to the breaches of the hypothesized norms, as opposed to some other aspects of the storyboards. When compared to the reactions to the breach storyboards, the reactions to the control storyboards were in general higher in terms of their overall and segment-specific *appropriateness* ratings. They also contained higher numbers of descriptive, as opposed to evaluative, statements in the open response fields. Furthermore, there were roughly equal numbers of positive and negative statements of attitude throughout the responses to the control versions of the storyboards. This was not the case for the breach storyboards, for which there were significantly more negative evaluative statements.

The results of the analysis of the responses to the rating questions and the open responses questions indicate that the control storyboards were reasonable representations of routine or natural teaching. Some participants liked what they saw, some took issue with what they saw, but in the aggregate, the reactions were relatively muted (as evidenced by the higher scores for the appropriateness ratings) and diverse (in terms of the mixture of positive and negative statements of attitude that were evident in the open responses). By contrast, the breach storyboards represented teaching that was rated low and that yielded negative statements of attitude. The matched breach/control design of the storyboards provides a warrant that this response pattern was caused by the breaches of the norm, rather than by some other aspect of the storyboards. Furthermore, there is evidence that participants' negative reactions to the departures from the norm were such because of the unexpected, rather than unreasonable, nature of the teacher's departure from the hypothesized norm.

This is evident in some of the open responses to the breach versions of the storyboards. The storyboards were designed for the breaches to represent reasonable alternatives to the hypothesized-to-be-normative classroom practices. Throughout the corpus, there were examples of responses that indicated that the teachers' actions in the breach versions of the storyboards could have desirable outcomes, such as demonstrating for students that diagrams should not be taken for granted, or emphasizing the importance of labeling a figure completely before using those labels in the proof. Responses that explicitly cast the teacher's breach of the hypothesized norm in such a positive light were by no means abundant in the corpus of open responses, yet the presence of even a few of these responses indicates that the teachers' actions to breach the norm were viewed as warranted by at least some participants in the study.

Significance of this study

The research reported here has scholarly and pedagogical significance. The principal knowledge claims of the study are that there are routine ways that semiotic resources are used when doing proofs in geometry and that secondary teachers recognize these routines. Both results have significance for mathematics education as a field.

The articulation of the *channel, details,* and *sequence* norms provide finer-grained descriptions of the semiotic terms of the exchange of proofs for knowledge claims when doing proofs in geometry. Scholars have argued that mathematics classrooms primarily develop literacy with written mathematics (O'Halloran, 2005; Kress & van Leeuwen, 2006). Such claims about the role of writing in mathematics classrooms tend to be general claims that are linked to historical biases in the discipline of mathematics against non-symbolic modes of communication.

The work conducted for this dissertation contributes specific accounts of how semiotic resources are used in geometry classrooms. The multimodal analysis of video episodes of students and teachers presenting and checking proofs allowed me to make specific hypotheses about the expectations teachers have for how students use semiotic resources. The virtual breaching experiment allowed me to test the extent to which specific hypotheses about the normative ways of using semiotic resources were recognizable by secondary mathematics teachers. The combination of qualitative analysis to identify semiotic norms and an experiment to measure the extent to which teachers recognized the the hypothesized norms provides specific evidence of the limited ways that non-written modes of communication are used in geometry classrooms.

The results of the study are also pedagogically significant. Improving the teaching of proof across the grades has long been a focus of efforts to make mathematics education more effective. The research on proof has tended to approach the problem of teaching proof more effectively as a challenge of diagnosing and repairing conceptual or psychological deficiencies about proof that are held by students or teachers. The instructional exchanges perspective adopted for this study provides a different way forward.

The multimodal disparities between the disciplinary communication practice of *chalk talk* and student presentations of proof in geometry classrooms indicate opportunities for teachers to create value for non-written semiotic resources when students are presenting proofs. Rather than focus exclusively on teaching students how to write proofs, teachers could create opportunities for teaching students how to generate proofs—using speaking, writing, drawing, and gesturing—while providing commentary on what they are developing as it is developed . Creating occasions for students to develop their multimodal communication skills could help students hone an area of authentic mathematical practice that is often overlooked in mathematics classrooms.

The creation of opportunities for students to develop proof presentation skills that are in line with those used by experts during *chalk talk* will require creating tasks for students that naturally lead to multimodal exchanges. This is how the theory of instructional exchanges can be used as a lever for affecting practice. Once the knowledge at stake has been identified, the challenge is to engineer a task that will provide an opportunity for the student to develop or draw on that knowledge in order to complete it.

In the area of helping students develop more authentic proof practice, Cirillo and Herbst (2012) provide a model of this reverse engineering. They identified the kinds disciplinary proof practices that students could be taught to develop, and then created tasks that could specifically target those practices. The research conducted here suggests that there are opportunities in geometry classrooms for students to hone their skills in non-written modes of mathematical communication.

To encourage students to develop fluency with non-written modes of communication, teachers could ask students to present proofs to the class verbally and without writing the statements and reasons for the proof on the board. If the goal of the task was for the students to present the proof in a way that the student's classmates could follow, this could lead the student toward a consideration of how to effectively coordinate speaking while gesturing at a diagram during the presentation of a proof. A task such as this would emphasize the speaking and gestural modes of mathematical communication and could create an opportunity for students to develop the capacity to use these modes.

A different example of a task that could help students develop their literacy in other modes of communicating about proof could be a task that requires students to present a proof to the class using a diagram that does not have any labels. The target competence for such a task would be for students to coordinate conceptual language and pointing (and potentially other) gestures while formulating a mathematical argument. Chen and Herbst (2013) found that student gesture production is affected by the features of diagrams and that students used metaphoric and iconic gestures when interacting with diagrams that did not have labels. That teachers of geometry could engineer tasks that might specifically help students develop their fluency in non-written modes of mathematical communication is an opportunity for teachers in geometry classrooms. The research reported in this dissertation supports such efforts by providing an account of the possible terms in such multimodal exchanges.

Conclusion

For as long as mathematics has been taught in US public schools, there have been initiatives that have attempted to improve the quality of mathematics teaching in classrooms. A fundamental challenge for such initiatives is the paradox of *change without difference*: reform efforts that, in principle, could bring about fundamental shifts in classrooms emerge, in practice, as "shadows of their original intent" (Woodbury & Gess-Newsome, p. 763). Why might that happen? The patterns of classroom interaction that practicing teachers have honed through years of experience are robust. Initiatives that aim to effect change in the way that mathematics is taught need to contend with the realities of the already established practice of mathematics teachers. If we accept that premise, studies like this one become vital background information for designing reforms.

Descriptive studies of teaching, such as this dissertation, take for granted that the normative teaching that occurs in classrooms results from a state of equilibrium among competing influences (Herbst & Chazan, 2012). The teaching that occurs in classrooms has to be seen not judgmentally, that is, not as *a priori* deficient, but rather as a state of affairs that can be scientifically studied, described and eventually explained. The study reported here shows that it is possible to generate specific descriptions of teaching without analyzing a large, representative sample of classroom episodes. The method of hypothesizing norms and experimentally testing whether norms are recognized by teachers is a technique for investigating the degree to which local regularities of particular classrooms (such as those based on a review of small sample of classroom videos) are actually part of the shared practice of mathematics teaching.

The study reported here generated specific descriptions of practice that are meaningful to teachers. Such accounts of the practice of mathematics teaching provide a foundation for designing instructional interventions that are aligned with the realities and constraints of existing practice (Cobb, Zhao, & Dean, 2009). Through specific descriptions of teaching practice, there is an opportunity to meet teachers on their terms and plan classroom interventions that could bring changes *with difference* to mathematics classrooms.

REFERENCES

Aaron, W.R., Erickson, A.W., Dimmel, J.K., & Herbst, P.G. (2013). *A linguistically grounded coded scheme for open-ended responses.* Poster presented at the National Council of Teacher's of Mathematics Research Conference, Denver, Colorado.

Aaron, W. R., & Herbst, P. (2012). Instructional identities of geometry students. *The Journal of Mathematical Behavior*, *31*(3), 382-400.

Alibali, M. W., Young, A. G., Crooks, N. M., Yeo, A., Wolfgram, M. S., Ledesma, I. M., & Knuth, E. J. (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture*, *13*(2), 210-233.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science*, *15*(2), 106-111.

Arsac, G., Balacheff, N., & Mante, M. (1992). Teacher's role and reproducibility of didactical situations. *Educational studies in mathematics*, *23*(1), 5-29.

Artemeva, N., & Fox, J. (2011). The writing's on the board: the global and the local in teaching undergraduate mathematics through chalk talk. *Written Communication*, *28(4)*, 345–379.

Atkinson, P. (1988). Ethnomethodology: A critical review. *Annual review of sociology, 14*, 441-465.

Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In Bishop, A.J, Mellin-Olsen, S, & Van Dormolen, J (Eds.) *Mathematical knowledge: Its growth through teaching* (pp. 173-192). Netherlands: Springer.

Balacheff, N. (2008). The role of the researcher's epistemology in mathematics education: an essay on the case of proof. *ZDM*, *40*(3), 501-512.

Barwise, J., & Etchemendy, J. (1996). Visual information and valid reasoning. In G. Allwein & J. Barwise (Eds.), Logical reasoning with diagrams (pp. 3–25). New York, NY: Oxford University Press.

Bieda, K. N. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities. *Journal for Research in Mathematics Education 41(4)*, 351-382.

Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in medicine*, *21*(9), 1331-1335.

Bourdieu, P. (1998). *Practical reason: On the theory of action*. Stanford, CA: Stanford University Press.

Brousseau, G. (1997). Foundations and methods of didactique. In N. Balacheff, M. Cooper, R. Sutherland, & V.Warfield (Eds. & Trans.), Theory of didactical situations in mathematics: Didactique des Mathematiques 1970–1990 (pp. 21–75). Dordrecht, The Netherlands: Kluwer. (Original work published 1986)

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational studies in mathematics*, *24*(4), 359-387.

Chazan, D., & Herbst, P. (2012). Animations of classroom interaction: Expanding the boundaries of video records of practice. *Teachers College Record*, *114*(3), 1-34.

Chen, C. L., & Herbst, P. (2013). The interplay among gestures, discourse, and diagrams in students' geometrical reasoning. *Educational Studies in Mathematics*, *83*(2), 285-307.

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.

Cirillo, M., & Herbst, P. G. (2012). Moving toward More Authentic Proof Practices in Geometry. *Mathematics Educator*, *21*(2), 11-33.

Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting design experiments to support teachers' learning: A reflection from the field. *The Journal of the Learning Sciences*, *18*(2), 165-199.

Cohen, L., Manion. L., & Morrison, K.,(2007). *Research Methods in Education.* New York, NY: Routledge.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational evaluation and policy analysis*, *25*(2), 119-142.

Davis, P. J. (1972). Fidelity in mathematical discourse: Is one and one really two?. *American Mathematical Monthly*, 252-263.

Davis, J., & Rebelsky, S. A. (2007). Food-first computer science: starting the first course right with PB&J. *ACM SIGCSE Bulletin*, *39*(1), 372-376.

Dimmel, J.K., & Herbst, P.G. (2015). The semiotic structure of geometry diagrams: How textbook diagrams convey meaning. *Journal for Research in Mathematics education. 46(2)*.

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist*, *23*(2), 167-180.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational studies in mathematics*, *38*(1-3), 85-109.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International journal of public health*, 1-6.

Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders*, *32*(1), 19-34.

Feynman, R. P. (1999). *The pleasure of finding things out: The best short works of Richard P. Feynman*. New York, NY: Basic Books.

Garfinkel, H. (1963). A conception of and experiments with" trust" as a condition of concerted stable actions. The production of reality: Essays and readings on social interaction, 381-392.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in cognitive sciences*, *9*(5), 234-241.

González, G., & Herbst, P. G. (2006). Competing arguments for the geometry course: Why were American high school students supposed to study geometry in the twentieth century?. *International Journal for the History of Mathematics Education*, *1*(1).

Gregory Jr, S. W. (1982). Accounts as Assembled from Breaching Experiments. *Symbolic Interaction*, *5*(1), 49-63.

Greiffenhagen, C. (2014). The materiality of mathematics: Presenting mathematics at the blackboard. *The British journal of sociology, 65(3),* 502-529.

Greiffenhagen, C., & Sharrock, W. (2011). Does mathematics look certain in the front, but fallible in the back?. *Social Studies of Science*, *41*(6), 839-866.

Halliday, M., & Matthiessen, C. (2004). *An introduction to functional grammar*. New York, NY: Routledge.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange*, *21*(1), 6-13.

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational studies in mathematics*, *44*(1-2), 5-23.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *Research in collegiate mathematics education III*, *7*, 234-282.

Herbst, P. G. (2002). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education*, 176-203.

Herbst, P. G. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics*, *49*(3), 283-312.

Herbst, P. G. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education*, 313-347.

Herbst, P., Kosko, K. W., & Dimmel, J. K. (2013). How are geometric proof problems presented? Conceptualizing and measuring teachers' recognition of the diagrammatic register. In M. V. Martinez & A. C. Superfine (Eds.), *Proceedings of the 35th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 179–186). Chicago, IL: University of Illinois at Chicago.

Herbst, P., & Balacheff, N. (2009). Proving and knowing in public: What counts as proof in a classroom. In Stylianou, D.A., Blanton, M.L., & Knuth, E.J. (Eds.) *Teaching and learning proofs across the Grades: A K-16 perspective*, (pp. 40-70). New York, NY: Routledge.

Herbst, P., & Brach, C. (2006). Proving and doing proofs in high school geometry classes: What is it that is going on for students?. *Cognition and Instruction*, *24*(1), 73-122.

Herbst, P., & Chazan, D. (2003). Exploring the practical rationality of mathematics teaching through conversations about videotaped episodes: The case of engaging students in proving. *For the learning of Mathematics*, 2-14.

Herbst, P., & Chazan, D. (2006). Producing a viable story of geometry instruction: What kind of representation calls forth teachers' practical rationality. In *Proc. of the 28 th PME-NA Conference* (Vol. 2, pp. 213-220).

Herbst, P., & Chazan, D. (2011). Research on Practical Rationality: Studying the justification of actions in mathematics teaching. *The Mathematics Enthusiast, 8(3),* 405-462.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM*, *44*(5), 601-612.

Herbst, P., Chazan, D., Chen, C. L., Chieu, V. M., & Weiss, M. (2011). Using comics-based representations of teaching, and technology, to bring practice to teacher education courses. *ZDM*, *43*(1), 91-103.

Herbst, P.G., & Dimmel, J.K. (2011). *Teaching geometry with problems and its demands of knowledge management.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Herbst, P., & Kosko, K. W. (2014). 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*, *17*(*6*)*,* 515-537.

Herbst, P.G., Kosko, K.W., & Dimmel, J.K. (2013). How are proof problems presented? Conceptualizing and measuring teachers' recognition of the diagrammatic register. In Martinez, M., & Castro Superfine, A (Eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 179-186). Chicago, IL: University of Illinois at Chicago.

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.

Herbst, P., Chen, C., Weiss, M, & González, G., with Nachlieli, T., Hamlin, M., & Brach, C. (2009). "Doing Proofs" in geometry classrooms. In Stylianou, D.A., Blanton, M.L., & Knuth, E.J. (Eds.) *Teaching and learning proofs across the Grades: A K-16 perspective*, (pp. 250-268). New York, NY: Routledge.

Hersh, R. (1991). Mathematics has a front and a back. *Synthese*, *88*(2), 127-133.

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics*, *24*(4), 389-399.

Hirsh, I. J., & Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of experimental psychology*, *62*(5), 423.

Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, *31*(8), 717-724.

Jacobs, J. K., & Morita, E. (2002). Japanese and American teachers' evaluations of videotaped mathematics lessons. *Journal for Research in Mathematics Education*, 154-175.

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. London: Routledge.

Kendon, A. (1994). Do gestures communicate? A review. *Research on language and social interaction*, *27*(3), 175-200.

Kilpatrick, J. (2001). Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics*, *47*(1), 101-116.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for research in mathematics education*, *33(5),* 379-405.

Kress, G. (2003). *Literacy in the new media age*. New York, NY: Routledge.

Kress, G. (2009).What is mode? In Jewitt, C. (Ed.) *The Routledge handbook of multimodal analysis*, 54-67. London: Routledge.

Kress, G., & Van Leeuwen, T. (2006). Reading images: The grammar of visual images. Oxon, Oxford: Routledge.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.

Lemke, J.L. (2013) Multimedia and discourse analysis. In Gee, J. P., & Handford, M. (Eds.). *The Routledge handbook of discourse analysis,* 79-89. New York, NY: Routledge.

Livingston, E. (1999). Cultures of proving. *Social Studies of Science*, *29*(6), 867-888.

Lockhart, P. (2009). *A mathematician's lament*. New York: Bellevue Literary Press.

Manders, K. (2008). The Euclidean Diagram. In Mancosu, P. (Ed.) *The philosophy of mathematical practice* (pp. 112–183). Oxford: Oxford University Press.

Mariotti, M. A. (2006). Proof and proving in mathematics education. *Handbook of research on the psychology of mathematics education: Past, present and future*, 173-204.

Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. New York, NY: Continuum.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education 20(1)*, 41-51.

Martin, J. &P. White. 2005. The Language of Evaluation: Appraisal in English. New York, NY: Palgrave Macmillan

McNeill, D. (2008). *Gesture and thought*. Chicago: University of Chicago Press.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Mehan, H., & Wood, H. (1975). *The reality of ethnomethodology*. New York: Wiley.

Morgan, C., Craig, T., Schuette, M., & Wagner, D. (2014). Language and communication in mathematics education: an overview of research in the field. *ZDM*, *46*(6), 843-853.

Nachlieli, T., Herbst, P., & González, G. (2009). Seeing a colleague encourage a student to make an assumption while proving: What teachers put in play when casting an episode of instruction. *Journal for Research in Mathematics Education 40(4)*, 427-459.

Nolan, K. (2012). Dispositions in the field: Viewing mathematics teacher education through the lens of Bourdieu's social field theory. *Educational studies in mathematics*, *80*(1-2), 201-215.

Núñez, R. E. (2009). Gesture, inscriptions, and abstraction: The embodied nature of mathematics or Why mathematics education shouldn't leave the math untouched. In Roth, W.M. (Ed.) *Mathematical representation at the interface of body and culture*, 309-328. Charlotte, NC: Information Age Publishing.

O'Donnell, M. (2008, April). The UAM CorpusTool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA, Almeria, Spain* (pp. 3-5).

O'Halloran, K. (2005). *Mathematical discourse: Language, symbolism and visual images*. New York, NY: Continuum.

O'Halloran, K. (2011). The semantic hyperspace: Accumulating mathematical knowledge across semiotic resources and modalities. *Disciplinarity: Functional linguistic and sociological perspectives*, 217-236. New York, NY: Continuum.

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* Thousand Oaks, CA: SAGE Publications, inc.

Rafalovich, A. (2006). Making sociology relevant: The assignment and application of breaching experiments. *Teaching sociology*, 156-163.

Rav, Y. (1999). Why do we prove theorems?. *Philosophia mathematica*, *7*(1), 5-41.

Read, J., & Carroll, J. (2012). Annotating expressions of appraisal in English. *Language resources and evaluation*, *46*(3), 421-447.

Rota, G. C. (1997). The phenomenology of mathematical beauty. *Synthese*, *111*(2), 171-182.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of 'well-taught' mathematics courses. *Educational psychologist*, *23*(2), 145-166.

Sears, R., & Chávez, Ó. (2014). Opportunities to engage with proof: the nature of proof tasks in two geometry textbooks and its influence on enacted lessons. *ZDM*, *46*(5), 767-780.

Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction*, *16*(3), 228-240.

Sfard, A. (2008). *Thinking as communicating: Human Development, the Growth of Discourses, and Mathematizing (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Cengage.

Sierpinska, A. (2005). Discoursing mathematics away. In Kilpatrick, J., Hoyles, C., & Skovsmose, O. (Eds.) *Meaning in mathematics education* (pp. 205-230). New York, NY: Springer.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50-80.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap.* New York, NY: Free Press.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). The TIMSS videotape classroom study. *Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* Washington, DC: US Government Printing Office.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education 38(3)*, 289-321.

Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education 40(3)*, 314-352.

Tappenden, J. (2008). Mathematical concepts: Fruitfulness and naturalness. In Mancosu, P. (Ed.) *The philosophy of mathematical practice* (pp. 276-301). Oxford: Oxford University Press.

Thurston, W. P. (1994). On proof and progress in mathematics. *arXiv preprint math/9404236*.

van Leeuwen, T. (2005). *Introducing social semiotics*. London, England: Routledge.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516-551.

Weber, K. (2004). Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *The Journal of Mathematical Behavior*, *23*(2), 115-133.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education 39(4)*, 431-459.

Weiss, M., & Herbst, P. (2007, April). *“Every single little proof they do, you could call it a theorem”: Translation between abstract concepts and concrete objects in the geometry classroom.* Paper presented at the Annual Meeting of American Educational Research Association, Chicago, IL.

Woodbury, S., & Gess-Newsome, J. (2002). Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. *Educational Policy*, *16*(5), 763-782.

Yore, L. D., Pimm, D., & Tuan, H. L. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education*, *5*(4), 559-589.

Yore, L. D., Pimm, D., & Tuan, H. L. (2007). The literacy component of mathematical and scientific literacy. International Journal of Science and Mathematics Education, 5(4), 559-589.