An Investigation of the Relationship Between Two Norms of the Instructional Situation of Geometric Calculation with Algebra in U.S. High School Geometry

by

Nicolas Boileau

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2021

Doctoral Committee:

Professor Patricio G. Herbst, Chair
Professor Deborah L. Ball
Professor Richard Gonzalez
Professor Edward A. Silver

Nicolas Boileau

nboilea@umich.edu

ORCID iD: 0000-0001-5314-7493

Dedication

This dissertation is dedicated to Elizabeth Tuscano, as without her love and support it would not have been possible.

Acknowledgements

When I came to Michigan, I had a keen interest in mathematics education research, but I lacked much of the knowledge and skills that it requires. To the extent that I have sense developed these things, I am indebted to my advisor and chair of my dissertation committee, Patricio Herbst. Through his mentorship and trust in my ability to manage various types of work in his lab during my time at Michigan, I learned a great deal about research, in particular in mathematics education. I am similarly indebted to all of the other members of his lab who I have worked with and who have supported me: Chieu Vu Minh, Amanda Milewski, Justin Dimmel, Ander Erickson, Rachel Snider, Annick Rougee, Kristi Hanby, Inah Ko, Umut Gürsel, Emanuele Bardelli, Mollee Shultz, Michael Ion, Claudine Margolis, Scott Bridges, Irma Stevens, Enes Akbuga, Heather Bleecker, and Inese Berzina-Pitcher, as well as a number of masters and undergraduate students that I've had a chance to work with in this context. Similar thanks to colleagues at University of Maryland that I've work with on projects led by Pat and his long-time colleague Daniel Chazan: Orly Buchbinder, Elizabeth Fleming, Eileen Drusjack, and of course Dan himself.

Next, thanks to the other members of my dissertation committee: Deborah Ball, Edward Silver, and Richard Gonzalez. From their willingness to stay on the committee when I changed projects to their feedback during the subsequent proposal defense, drafts of my dissertation, and presentations of the work, I learned about both mentorship and research. They have each pushed my thinking in many important ways, in particular, by helping me think about how I could make my work more accessible and compelling to researchers both inside and outside of the field of education.

I am also forever indebted to my earlier academic mentors at Concordia University: Anna Sierpinska and Nadia Hardy. To Anna, for taking a chance on me and for the thoughtful and caring mentorship that she offered as my advisor. To Nadia, for the mentorship she offered as the P.I. on the first research project that I ever worked on. To both of them, for suggesting that I pursue a Ph.D. in mathematics education. And to both of them, as well as the other instructors of the courses that I took while in the M.T.M program, for introducing me to the field.

Back to Michigan: I am similarly thankful to the many brilliant scholars who taught the courses I took at the School of Education. I am thankful for the university's various educational programs and centers, such as the Inter-university Consortium for Political and Social Research (ICPSR) Summer Program in Quantitative Methods of Social Research and the center for Consulting for Statistics, Computing and Analytics Research (CSCAR). Special thanks to Corey Powell, Josh Errickson, and Michael Clark—consultants at CSCAR that I met with regularly during my time here.

I also owe a great deal of thanks to the community of graduate students who supported me. They include my math education cohort—Charles Wilkes II, Amber Willis, Kwame Yankson, Carolyn Masserang, and Jill Mortimer—as well as a group of scholars that I worked with each week, while dissertating: Naomi Mae W., Gordon Palmer, Christina Cross, Asya Harrison, Channing Mathews, Darrius Robinson, Sarah McKellar, and Rebecca Gadd. I am thankful to them not only for the various ways that they supported my work, but also for the love they've shown me. I am similarly thankful to many other students at Michigan that welcomed me into their lives: Will Toledo, Max Altman, Anne Cawley, Tonya Kneff, Chauncey Smith, Dana

Nickson, Ebony Perouse-Harvey, Christina Morton, Kim Ransom, Crystal Wise, and Jeff Grim, to name a few.

To my family back home—my dad, my mom, my brothers, cousins, uncles and aunts, and grandparents; Sam and Marty, and their families—and to the friends that I've had since high school, as well as their families: thank you. One of the great blessings in my life is to have had such community, to be able to consider such a large group of people family. I am particularly thankful for them now, as I write these acknowledgements and reflect on the roles they've all played in getting me here.

Last, but most importantly, to my wife, Liza. Her constant love is what got me through this. During the toughest times of this process, she tended to me. She picked up my slack when I fell behind on chores. She gave me hugs. She made me laugh, every day. She convinced me I was capable when she knew I didn't believe it. She went for walks with me when I needed a break from the work. She helped me think through some tough decisions (e.g., related to data analysis and writing). She even copyedited my dissertation! She encouraged me to apply to Ph.D. programs, then moved to Windsor, then to Ann Arbor, all so that I could pursue this dream. I have no better example of selflessness or love than her and these acts. And it is for these reasons and more in that I dedicate this dissertation to her.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Studies have demonstrated that norms have considerable influence on human behaviour, in particular, that of teachers and students in mathematics classrooms. Studies have also shown that breaches of norms are frequently sanctioned, sometimes positively, but typically negatively. The present study builds on that literature by investigating two other potential consequences of breaching norms of mathematics instruction: that breaches of one norm of a given instructional situation may lead teachers to abandon their expectations that other norms of that situation will be followed and/or alter their attitudes towards breaches of those norms.

I focus on the relationship between two hypothesized norms of geometric calculations with algebra (GCA) in U.S. high school geometry. One of them, the GCA-Figure norm, stipulates that the GCA problems that U.S. high school geometry teachers assign are expected to have geometrically-meaningful solutions. The other, the GCA-Theorem norm, stipulates that, when solving GCA problems, students are expected to document their algebraic work, to occasionally verbally state the geometric properties that warrant the equations that they set up, but not to document those properties.

To confirm the existence of those norms and investigate whether breaches of the GCA-Figure norm would have either of the aforementioned consequences, I conducted a virtual breaching experiment. This consisted of randomly assigning U.S. high school mathematics teachers to one of three sets of multimedia questionnaires. Each questionnaire confronted the participant with a storyboard representation of a classroom scenario in which each of the two norms is either breached or followed. Their reactions to each storyboard were captured through a

set of open- and closed-response items. Scores, based on coded open-responses and closed-responses, were compared across experimental conditions, using statistical models. This was done to predict whether experienced geometry teachers would be more likely to recognize decisions to breach either norm than decisions to follow it (evidence that the hypothesized norm exists), to deem decisions to breach it more acceptable than decisions to follow it, and/or to remark or disapprove of decisions to breach the GCA-Theorem norm when the GCA-Figure norm is followed than when it is breached.

Results suggest that experienced geometry teachers' expectations of GCA problems are well-represented by the above statement of the GCA-Figure norm, but that their expectations of solutions to GCA problems are slightly different than hypothesized. Namely, they suggest that experienced geometry teachers expect students to document their algebraic work, but not to share their geometric reasoning (verbally or in writing). In terms of attitudes towards breaches, results suggest that experienced geometry teachers are generally opposed to problems that breach the GCA-Figure norm, but do not provide much information about their attitudes towards students sharing their geometric reasoning, suggesting the need to develop alternate ways of measuring such attitudes in future research. Lastly, results suggest that experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm are not dependent on whether the GCA-Figure norm is followed, but that such teachers may abandon their expectation that the GCA-Theorem norm will be followed when the GCA-Figure norm is breached.

While the dissertation's main contribution is to our understanding of norms of mathematics instruction, it also has implications for instructional improvement. Namely, the latter result suggests that changing even a very specific behaviour may alter whole systems of expectations—something that reformers must consider when anticipating what their recommendations will require.

Chapter 1 - Introduction

Norms—(typically implicit) behavioural rules that people expect each other to follow in situations of a given type—are widely accepted as having considerable influence on human behaviour (Legros & Cislaghi, 2020). In fact, their influence is so strong that many have argued that descriptions of norms can be used to characterize the situations to which they are thought to apply (e.g., Goffman, 1964). For the same reason, some have suggested that an understanding of norms is essential to social change, claiming that efforts to change social behaviour that do not alter people's expectations of what others will do, or assuage their fears that breaches of norms will be punished, are unlikely to be successful (e.g., Bicchieri, 2017).

As a mathematics education researcher interested in understanding and describing how mathematics is taught and learned in schools (e.g., in a particular course of study)[1] and who believes that proposed instructional changes should be rooted in a deep understanding of why teachers and students do the things that they do, I have spent the past several years trying to understand how norms influence the decisions of teachers and students in mathematics classrooms. In the present study, building on earlier research that showed that breaches of norms[2] are often sanctioned (sometimes positively, but typically negatively) (e.g., Garfinkel, 1963; Dimmel & Herbst, 2017), I investigate two other potential consequences of breaching norms of mathematics instruction: that breaches of one norm of a given instructional situation may lead teachers to abandon their expectations that other norms of that situation will be

[1] I.e., how teachers and students interact around the content that students are expected to learn in such environments (Cohen, Raudenbush, & Ball, 2003).

[2] I.e., actions that violate norms.

followed and/or alter their attitudes towards breaches of those norms.[3] As this question is quite broad, I focus on a specific case: the relationship between two norms of a recurrent instructional situation in a particular course of studies, namely, the instructional situation of geometric calculations with algebra (GCA) in U.S. high school geometry. In this chapter, I explain and justify the choice to focus on this context—course of studies and situation within it—as well as the norms that I investigated.

1.1. An Argument for Focusing on Courses of Study, in Particular, U.S. High School Geometry

While it is quite reasonable to expect that certain patterns of behavior would exist across multiple courses of study—such as students engaging in recommended mathematical practices (e.g., those proposed by the National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) or teachers following recommendations on how to support them in doing so (e.g., those proposed by the National Council of Teachers of Mathematics, 2014)—it is also reasonable to expect that the teaching and learning of mathematics differs importantly across courses of study. One reason to expect that is that their curriculums (i.e., the content to be taught and learned) are different. For that reason, some have argued for the importance of developing subject-specific accounts of mathematics instruction that demonstrate such differences (e.g., Herbst & Chazan, 2017).

[3] While this question about norms of mathematics instruction is still open, theories of norms outside of education provide some precedent. For example, Bicchieri (2006, p.56) suggests that norms are activated in the mind of an individual when they frame the situation they are in as one to which those norms apply. Similarly, she suggests that, as we continue to monitor our environments, we may notice things that have us reconsider whether we correctly framed the situation and, possibly, reframe it. Therefore, when we reframe the situations we are in, we may abandon expectations activated by the original framing that we think no longer apply. For those more familiar with Schoenfeld's (2010) theory of decision-making, this idea is also implicit in that theory, as he posits that, when we recognize situations as familiar, cognitive scripts that prescribe how to act in those situations are activated. While Schoenfeld does not describe the role of norms in decision making, Bicchieri helps us see how his theory could be extended by suggesting that norms are embedded in scripts and that both norms and scripts are therefore activated when one recognizes a situation as familiar (and categorizes it as a situation to which those norms and scripts apply).

In terms of justifying the choice to focus here on U.S. high school geometry, specifically, there are a few reasons[4]. One of those is that geometry is one of the most commonly taught high school mathematics courses in the U.S.—in fact, it is one of the few mathematics courses required to graduate from high school in many states (Macdonald, Zinth, & Pompelia, 2019). But perhaps a more compelling reason is that it offers many unique opportunities to learn theoretical mathematics. For example, it is in U.S. high school geometry that students typically learn how to read and write mathematical proofs (Otten, Males, & Gilbertson, 2014), which is often considered the primary activity of professional mathematicians. In high school geometry, students also learn to reason in particular ways with diagrams (Dimmel & Herbst, 2015; Hsu, 2010a). Along those lines, in section 1.4, I describe the opportunities to learn to reason with diagrams, and potentially to prove, offered by the instructional situation of GCA. But, first, I define the notion of an instructional situation.

1.2. An Argument for Focusing on Instructional Situations

The term *instructional situation* has been used casually (without definition) by several researchers (e.g., Cobb, Wood, Yackel, & McNeal, 1992; Kuntze, 2011). The meaning that I believe those authors intended to convey, based on my reading of those papers, is that an instructional situation is any moment in a classroom when a teacher is working with their students towards an instructional goal. However, throughout the dissertation, I draw on Herbst (2006) notion of an *instructional situation*[5], which was inspired by Erving Goffman's notion of a *social situation*. Goffman (1964, p.231) defined that concept as follows:

[4] For a more-thorough discussion of this topic, see González & Herbst (2006).

[5] That is, the term *instructional situation* was first formally defined in that paper. It was used casually in prior literature (e.g., Cobb, Wood, Yackel, & McNeal, 1992), seemingly (based on my understanding of the sentences in which it was used) to refer to any moment in a classroom when a teacher is working with their students towards an instructional goal.

> I would define a social situation as an environment of mutual monitoring possibilities, anywhere within which an individual will find himself accessible to the naked senses of all others who are 'present,' and similarly find them accessible to him. According to this definition, a social situation arises whenever two or more individuals find themselves in one another's immediate presence, and it lasts until the next-to-last person leaves.

As such, Goffman argued that social situations "constitute a reality" that frame how we interpret the actions of other people and therefore "warrant analysis in their own right" (ibid.). Along those lines, he argued that one of their defining features were the "[c]ultural rules [that] establish how individuals are to conduct themselves by virtue of being [in the situation]" and that, "when adhered to, socially organize the behaviour of those in the situation" (ibid.)—in other words, *norms*. However, to apply this notion to the study of mathematics instruction, Herbst needed to answer the following question: what situations recur within mathematics classrooms with enough frequency that teachers (and students) would have developed expectations about them? The seed of an answer lied in Doyle's (1988) suggestion that most of the academic tasks that students are assigned (in mathematics and other classrooms) can be categorized as *familiar* (rather than *novel*). He defined the concept of *familiar tasks* as follows:

> Familiar [tasks] consists of routinized, recurring exercises, such as warm-ups and practice sets, in which memory or relatively standardized operations or algorithms are used to generate products.

> There is little ambiguity about what to do and how to do it and little risk that things will go wrong along the way. Such work creates only minimal demands for students to interpret situations [in which they are tasked with doing such work]…
>
> (Doyle, 1988, p.173)

In other words, according to Doyle, students recognize moments in classrooms when they are assigned familiar tasks as situations that have (often implicit[6]) rules (i.e., norms) that stipulate how they are to act in those situations. Herbst (2006) noted that teachers familiar with those situations would not only have expectations of student work, but also anticipations of how they could *exchange* that work for claims on the students' knowledge. Along those lines, Herbst and Chazan later defined instructional situations as follows:

> Conceptually, an instructional situation is a type of encounter where an exchange can happen between (1) specific mathematical work done by students and their teacher in moment-to-moment interaction and (2) a claim on students' knowing of a specific item of knowledge at stake. Intuitively one could think of an instructional situation as including a mathematical task and the element of the curriculum that the completion of the task enables the teacher to lay claim on.[7]
>
> (Herbst & Chazan, 2011, p.412)

---

[6] How implicit norms are passed on is actually an open question, particularly in education, but arguably across all fields. However, outside of education, there are a number of theories, including the hypothesis that we seek to coordinate our behavior with others and fear negative sanctions that come with breaching norms. For an overview, see Bicchieri, Santuoso, & Muldoon (2011). In education, an example of this would be students mimicking the approaches that their teacher takes to solving particular types of problems, even when the teacher does not explicitly teach some or all of the method.

[7] Herbst's original definition was very similar: "any one of the customary ways in which classroom actions are framed into units of work so as to be traded in for (or accounted to) claims over the knowledge at stake" (Herbst, 2006, p.316). I provide the definition above instead simply because it is explicit about who is doing the work as well as who is inferring what students know on the basis of that work.

In terms of the norms that are activated in the mind of a teacher[8] when they frame work on a given task as an instance of a particular instructional situation, Herbst and Chazan (2011) go on to explain that the norms of instructional situations include both expectations about the work to be done by the teacher and by the students as well as expectations about the temporal order in which that work will be done. More specifically, they include (1) expectations about the work that students will produce[9], (2) expectations about the work that teachers (e.g., both they and other teachers of that course) should assign, (3) expectations about how teachers should interpret and respond to students' work, as well as (4) expectations about the sequence of both students' actions (e.g., the steps in their solutions) and teachers' actions (e.g., when they can intervene on work being presented at the board). For example, when *doing proofs* in U.S. high school geometry, teachers expect that students will produce a mathematical proof and that they should exchange that work for claims on students' understandings of the mathematical properties used to justify their claims as well as their abilities to construct such mathematical arguments, or "do proofs" (Herbst, 2010, p.54). The proof itself should be presented as a sequence of statements and reasons (i.e., justifications of those statements), organized into a two-column table, which begins with the premise(s) of the mathematical proposition to be proved and ends with its conclusion (Herbst, Chen, Weiss, & González, 2009).

[8] While there is evidence that students are aware of norms in other literature (e.g., Brousseau & Warfield, 1999; Coob, Wood, Yackel, & McNeal, 1992) Herbst and Chazan have focused their efforts on building a model of teacher decisions making (Herbst & Chazan, 2012; Chazan, Herbst, & Clark, 2016) and have not investigated whether students are aware of the norms of instructional situations that they have described.

[9] Arguably, the teacher's expectations about what the task could help them teach, or (equivalently) what it could help the students learn, may differ across the stages of task implementations—from when the task is represented in their instructional materials, to when it is set up by the teacher in the classroom, to when it is taken up by the students (Stein, Grover, & Henningsen, 1996).

There are several reasons to study norms of instructional situations. Here I provide two. The first is that, through studying and describing norms, we can gain and share (respectively) a clear picture of the current state of the teaching and learning of mathematics in schools[10]. This is because instructional situations account for the majority of students' opportunities to learn during their time in a given course of studies[11] (Herbst, 2010) and because instructional situations can be modelled by describing their norms (Herbst & Chazan, 2011). The second reason to study norms has to do with instructional improvement: Because norms have been shown to have a stronger influence on behaviour than personal beliefs (Bicchieri, 2017, p.10), efforts to reform current instructional practices that (explicitly or implicitly) aim to change teachers' personal beliefs (e.g., about students or mathematics) without changing their beliefs about what others expect them to do are likely to fail (in some sense or to some degree). But an understanding of how norms relate to social change also suggest an alternative approach: to investigate whether and how proposed changes require individuals to deviate from norms, so that one can decide whether justifications for breaches must be proposed or new norms must be established (for more on this, see, e.g., Bicchieri, 2017).

With the general topic of the present study (norms of instructional situations) now defined and justified, I can now define the instructional situation of GCA. After doing so, I describe the norms of GCA on which I focused in the present study.

---

[10] Of course, in doing so, they do not suggest that norms of instructional situations explain all of the variation that exists within those courses (let alone within non-standard courses or institutional settings). In fact, their theory of the practical rationality of mathematics teaching (Herbst & Chazan, 2012; Chazan, Herbst, & Clark, 2016) and the empirical work that supports it (Herbst, Chaan, Kosko, Dimmel, & Erickson, 2016) acknowledge several other covariates.

[11] Because most tasks assigned in U.S. schools, including their mathematics classrooms, are of familiar types (Doyle, 1988) and teachers' recognition of tasks as familiar (e.g., as a proof task) is what is hypothesized to cause them to frame the work as an instance of an instructional situation (e.g., as an instance of doing proofs).

1.3. The Instructional Situation of GCA

The instructional situation of GCA in U.S. high school geometry (Herbst, 2010) is one in which the student is presented with a diagram representing a geometrical figure (Herbst et al., 2017, p.59) in which the measures of certain elements of the figure (e.g., sides, angles, or arcs) are represented both pictorially (by strokes or spaces between them) and symbolically (by algebraic expressions); then the problem solver is asked to determine either the value of the variable(s) in those expressions or the measure of certain elements of the figure. These *GCA tasks* can be completed by setting up and solving one or more equations, on the basis of geometric properties of the figure. For example, the GCA problem in Figure 1.1 can be solved by equating the lengths of segments $\overline{AB}$ and $\overline{AC}$ ($2x+3 = 5x-12$), which is warranted by the definition of an isosceles triangle—a triangle with (at least) two congruent sides—and by the markings of the diagram indicating that $\overline{AB}$ and $\overline{AC}$ are those congruent sides. The problem solver could then solve that equation and use its solution to determine the length of segment $\overline{BC}$.

Determine the length of $\overline{BC}$

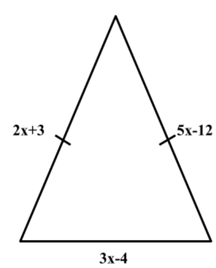


Figure 1.1: Sample GCA problem

From the perspective of the teacher, the student's work is evidence of their knowledge of relevant geometric properties, in particular those needed to set up the equation(s), as well as their knowledge of solving equations. As Herbst (2010) put it: "What is at stake [in this situation] is a claim on students' capacity to use a property they already know as well as a claim on maintaining knowledge of algebra skills" (p.53).

To better situate GCA tasks in the U.S. high school geometry curriculum, and thereby relate the mathematical topic of the present study to the mathematical topic of other studies, it is worth noting that GCA tasks are a subclass of a larger class of mathematical tasks called geometric calculation (GC) tasks. The latter also includes geometric calculation with number (GCN) tasks. Hsu (2010a) defined GC tasks as, "calculations done within mental or physical geometric diagrams on the basis of geometric principles or formulae" (p.16), and GCN tasks as, "*numerical* calculation done in relation to mental or physical geometric diagrams on the basis of geometric principles or formulae (e.g., calculating an angle measure in a triangle given that measures of the other two angles are 30° and 100°, respectively)" (p.2, *my emphasis*). She differentiated between GCN and GCA as follows:

> [I]f a GC task can be solved without using any algebraic skills to calculate measures, the GC is classified as a GCN no matter [if] algebraic expressions are used in the givens or not. On the other hand, if the algebraic skills are needed to calculate the unknowns, the GC task is categorized as a GCA.
>
> (Hsu, 2010a, p.18)

Her categorization of tasks is relevant for two reasons. As mentioned, one reason is that it helps define the mathematical topic of the present study—GCA. The other is that it helped me determine a larger, yet equally-relevant set of literature to review: Since the literature on GCA is thin, also reviewing literature on GCN allowed me to better explain how the present study contributes to the existing literature.

1.4. An Argument for Focusing on the Instructional Situation of GCA

One reason to study the instructional situation of GCA is that GCA problems are quite common in U.S. high school geometry textbooks[12][13]. As such, to the extent that textbooks represent the content taught in a given course, it is reasonable to expect that many of U.S. high school geometry students' opportunities to learn the content of that course consist of work on such problems. Yet not much is known about GCA problems, GCA tasks, or the instructional situation of GCA. For example, we do not know which characteristics of GCA problems (e.g., types of algebraic expressions used in a diagram) or student solutions are typically expected (i.e., normative).

The instructional situation of GCA is also an interesting context to examine because it requires students to reason with and across different mathematics representations (Behr, Lesh, Post, & Silver, 1983; Lesh, Post, and Behr 1987). Specifically, as mentioned above, GCA tasks ask students to calculate dimensions of a geometric figure, initially represented by pictorial aspects of a provided diagram (strokes and spaces) as well as by algebraic expressions. Their work therefore involved setting up one or more equations on the basis of geometric properties, solving the equation(s) using algebraic skills (learned in Algebra 1 and prior), then using the

[12] This claim is based on my perusal of popular U.S. high school geometry textbooks (e.g., Cummins et al., 2003). At this point, to my knowledge, there has not been any analyses of U.S. high school geometry textbooks that describe the frequency of GCA problems (or GC problems, more generally).

[13] In fact, there is some evidence that GC problems (more generally) are also common in other countries such as Australia (Stacey & MacGregor, 1999), England (Lang & Ruane, 1981), Germany (Schumann & Green, 2000), Taiwan (Hsu, 2010a), and Trinidad and Tobago (Hunte, 2018). That said, it should be noted that the norms that regulate work on GC problems in the U.S. might be different from the norms that regulate work on GC problems in any of those countries. For that reason, I am only including this comment as a footnote, as my intention is not to suggest that the results of the present study would necessarily pertain to mathematics teaching in other countries. However, the question of whether the norms of the instructional situation of GC (or any other instructional situation, in fact) are the same in two countries is an interesting empirical question.

solution(s) to evaluate those algebraic expressions. The fact that GCA tasks require reasoning with and across mathematical representations is relevant to motivating the present study as both mathematics education researchers and policy makers alike have been arguing for decades that students should have regular opportunities to engage in this mathematical practice (e.g., NCTM, 1989, 2000, 2018). There are at least two bases for the argument. First, some have claimed that considering different representations of a given concept is essential to developing a deep understanding of it: "using… different representations is like examining [a] concept through a variety of lenses, with each lens providing a different perspective that makes the picture (concept) richer and deeper" (Tripathi, 2008, p. 439). A second common argument is that professional mathematicians regularly and flexibly use mathematical representations to make sense of and solve mathematical problems, communicate their ideas, and read their colleagues' work (e.g., Schoenfeld, 1992; Stylianou, 2011) and that the work that students do in mathematics classrooms should resemble the work of professional mathematicians (e.g., Weiss, Herbst, & Chen, 2008).

Yet, it is also worth noting that the opportunity to reason with and across mathematical representations offered by GCA tasks could be rather unique, as they can be designed to require students to question the existence of the geometric figure purported to exist by the statement of the problem (including the diagram, when one is included). Specifically, whether or not the figure exists depends on the algebraic expressions to represent certain dimensions of the figure. For example, consider the GCA problem in Figure 1.2. Note that the pictorial components of this diagram (strokes and spaces) are the same as in the diagram in Figure 1.1. So are the algebraic expressions representing the lengths of segments $\overline{AB}$ and $\overline{AC}$. The only difference is the algebraic expression representing the length of segment $\overline{BC}$. To solve this problem, a student may again equate the lengths of segments $\overline{AB}$ and $\overline{AC}$ (2*x*+3 = 5*x*-12), find that *x*=5, then calculate the length of $\overline{BC}$ is 27. But this solution is impossible, given that the length of $\overline{AB}$ and of $\overline{AC}$ is 13, as this violates the *triangle inequality theorem* (which states that the sum of the lengths of any two sides of a triangle is less than the length of the third side). Therefore, the problem solver must conclude that the mathematical object that was purported to exist by the diagram—an isosceles triangle whose side lengths were related by the algebraic expressions used to represent them—actually does not exist. This type of reasoning is common in fields where diagrams (mental or physical) are used to model hypotheses (e.g., possible designs of structures), which sometimes prove to be incorrect.

Determine the measure length of $\overline{BC}$

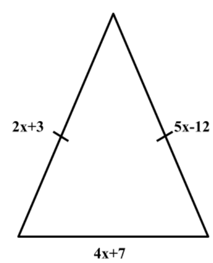


Figure 1.2: Another sample GCA problem

A final reason to study the instructional situation of GCA is that a theoretical argument can be made that work on GCA problems can be used to support students in developing some of the knowledge and skills needed to compose geometric proofs. This would be compelling to many mathematics educators as proof is arguably an even more iconic mathematical activity than reasoning across representations, yet one that students find particularly challenging (e.g., Reid & Knipping, 2010). It is reasonable to believe that this connection between GCA and geometric proof exists, because Hsu (2010a) empirically demonstrated that Taiwanese students' performance on GCN tasks was associated with their performance on geometric proof tasks and suggested that this was because they both require similar deductive reasoning with diagrams as well as knowledge of properties of geometric figures. For that reason and because completing GCN tasks is arguably less challenging than composing geometric proofs, she argued that GCN tasks could be used to support students in learning to compose geometric proofs. Since GCA tasks require the same geometric reasoning and knowledge as GCN tasks, it is reasonable to expect that GCA tasks could be used towards the same end.

1.5. The GCA-Figure and GCA-Theorem Norms

My investigation of the instructional situation of GCA began with a perusal[14] of popular U.S. high school geometry textbooks (e.g., Cummins et al., 2003) and observations of U.S. high school geometry lessons at local high schools (some of which I conducted for the purpose of this study and some of which I conducted while working as a field instructor). The goal of these activities was to determine patterns of behaviour that might be consequences of shared social expectations (i.e., norms). One thing that I noticed throughout both of these activities was that all of the GCA problems that I encountered were like the one represented in Figure 1.1, and unlike the one represented in Figure 1.2, in that none of them required students to question whether a figure of the type represented by the pictorial aspects of the diagram actually exists. Accordingly, I conjectured that GCA problems are expected to have geometrically meaningful solutions; that is, when someone sets up an equation on the basis of a property of the type of figure represented

[14] I hesitate to refer to this as a textbook analysis, as, while I looked at each GCA problem in each textbook that I considered, I was not systematic in how I interrogated each problem. That is because the goal of this portion of the work was simply to come up with conjectures that I would later have to confirm. In other words, while a systematic textbook analysis might have left me more confident in my claim that the trends that I observed are characteristic of the instructional situation of GCA (rather than the particular textbooks that I analyzed), it would not have been proof that they are the consequence of norms. Regardless of how systematic my review would have been, in order to confirm that the trends were indicative of norms would requires that one determine that those trends are (tacitly or explicitly) expected. And that was the goal of the study that I report in this dissertation.

by the pictorial components of the diagram, solves it, and uses its solution[15] to evaluate the algebraic expressions used to represent dimensions of the figure, it is expected that those calculations result in possible values for those dimensions—positive real numbers that satisfy all of the properties of the figure). This is what I came to call the *GCA-Figure norm*.

A second trend that I observed was that most GCA tasks in textbooks did not ask students to justify their work, in particular, the equations that they would set up. Similarly, when observing U.S. geometry classrooms, I noticed that students were rarely asked to state the geometric properties that warrant those equations and never asked to document them as part of their written work. Based on this, I conjectured that a second norm of the instructional situation of GCA was that, when U.S. high school geometry students work on GCA problems, they are expected to either only document their algebraic work, sometimes to verbally state the geometric properties that warrant the equations that they set up (e.g., when presenting their solution at the board), but never to write those properties. I came to call this the *GCA-Theorem norm*.

To provide an example of what it would look like for a student to follow the GCA-Theorem norm, imagine a student at the front of a class solving the problem in Figure 1.1 (or 1.2) and either simply writing the algebraic work underneath that figure or verbally stating the geometric property that warrants the equation ($2x+3 = 5x-12$): the definition of an isosceles triangle. Since people in social situations are expected to hold each other accountable to following norms (e.g., Garfinkel, 1963), I conjecture that it would likely also be deemed reasonable, and even go unnoticed, if the teacher were to ask the student to state that justification. In contrast, either a student's decision to write that property on the board or a

---

[15] I use the singular, *solution*, here because it was also my observation that students were almost always only required to set up and solve linear equations in one variable (the only exception I observed being a handful of GCA problems that required students to solve quadratic equations).

teacher's request for them to do so would likely draw attention and perhaps be sanctioned, as the student would be breaching the norm and the teacher would be asking them to do so. However, to be clear, in saying that, I am not suggesting that those decisions would be disapproved. For example, a teacher may think writing the property is important, especially when it was only recently introduced, as that could stimulate discussion, and deepen students' understanding, of the property. However, it is understandable that this would not be common, as my perusal of textbooks and classroom observations suggest that there is typically only one property that students would use to solve a given GCA problem (as in the examples above) and because students would typically be familiar with that property, as they would have been introduced to it prior and have used it to complete various other tasks (e.g., proofs, constructions, etc.).

1.6. An Argument for Focusing on These Two Norms?

As alluded to at the beginning of this chapter, I chose to investigate two norms of the instructional situation of GCA because I was interested in the question of whether breaching one norm of a given situation could alter teachers' and/or students' beliefs that other norms will be followed and/or their attitudes towards breaches of them. One reason I focused on these two norms is that, as mentioned earlier, there was clear value in breaching each of them (e.g., as breaching the GCA-Figure norm would provide students with the unique opportunity to reason about the existence of a mathematical object that is purported to exist and breaching the GCA-Theorem norm would require students to document and possibly defend their geometric reasoning as they need to when composing geometric proofs). Last, there was reason to think that the two norms may be related in that way: If it is rare as I observed for GCA problems to breach the GCA-Figure norm, both students and teachers (unless the breach was planned) may be disoriented by breaches of that norm and, consequently, reframe the situation as novel. In that case, there would be additional reason to have students document their reasoning: as the student would likely be surprised by the invalid side lengths that they produced, claims on prior knowledge of that geometric property would be all the teacher would be able to exchange for their work and, as mentioned above, these could be emphasized once documented. Moreover, an additional benefit of assigning problems that breach the GCA-Figure norm is that the new situation could be used as an opportunity to introduce students to *proof by contradiction*[16]—a type of proof that is often used by professional mathematicians, but one that U.S. high school students are rarely introduced to. Here, again, it would be important to document the property for students to be able to follow this (new type of) mathematical reasoning.

1.7. An Overview of the Present Study

While I expected that both teachers and (eventually) students expect the GCA-Figure and GCA-Theorem norm to be followed, in the study reported in this dissertation, I decided to investigate teachers' expectations of the instructional situation of GCA, as it is not yet understood when students become familiar with norms of instructional situations (and, therefore, when they become reliable informants on those norms). In contrast, there are several reasons to believe that a teacher of a given course would be familiar with its instructional situation (and their norms), especially after teaching the course a number of times (e.g., because they would have had to interpret instructional materials, possibly interacted with other teachers of the course, and may have even taken the course when they were students).

As such, the central research question of the present study is: are U.S. high school geometry teachers more likely to expect that the GCA-Theorem norm will be followed, or to

[16] In this case, that one can prove that a figure does not exist by assuming that it does, operating on the figure as one would if it did, and realizing that this results in a contradiction (e.g., that a side would have to have a negative length).

disapprove of breaches of the GCA-Theorem norm, when the GCA-Figure norm is followed (than when it is breached)? However, to answer this question, I needed to confirm that the above statements of the GCA-Figure and GCA-Theorem norms are in fact descriptions of teachers' expectations of the instructional situation of GCA. In order to accomplish both of these goals, I designed what is called a *virtual breaching experiment* (Herbst & Chazan, 2015). This consisted of administering a set of multimedia questionnaires that I designed to a national sample of high school mathematics teachers, most (303 of 480) of whom had substantial experience teaching high school geometry (three or more years) and could therefore be considered reliable informants on the norms of the instructional situation of GCA.[17] Each questionnaire presents the participant with a storyboard representation of a classroom scenario in which each of the hypothesized norms is either breached or followed (which I expected they would frame as an instance of the instructional situation of GCA, at least when the GCA-Figure norm is followed). The questionnaire also contains a set of items about each storyboard, in particular, about what the participant noticed and whether they think the teacher's actions were appropriate. As such, the experiment allowed me to answer the following three research questions:

1. Do the GCA-Figure and GCA-Theorem norms exist? That is,
    a. Do experienced geometry teachers expect GCA problems to have geometrically-meaningful solutions?

[17] The claim that, after three of more years of experience teaching a given course of studies, a teacher could be expected to be aware of the norms of its instructional situations is based on prior research on norms of instructional situations (e.g., Herbst, Aaron, Dimmel, & Erickson, 2013; Herbst, Nachlieli, & Chazan, 2011).

   b. Do experienced geometry teachers expect students to either only share their algebraic work, or to sometimes also verbally state the geometric properties that warrant the equations that they set up, but not to write those properties?

2. Do experienced geometry teachers deem decisions to breach either norm acceptable or at least as acceptable as decisions to follow them? That is,
   a. Do experienced geometry teachers deem other U.S. high school geometry teachers' decisions to assign problems that breach the GCA-Figure norm acceptable or at least as acceptable as decisions to assign problems that follow it?
   b. Do experienced geometry teachers deem students' decisions to breach the GCA-Theorem norm and/or U.S. high school geometry teachers' decisions to ask students to do so acceptable or at least as acceptable as decisions to follow it?
3. Are experienced geometry teachers more likely to expect that the GCA-Theorem norm will be followed, or to disapprove of breaches of the GCA-Theorem norm, when the GCA-Figure norm is followed (than when it is breached)?
   a. Are experienced geometry teachers' expectations about whether and how students will share the geometric properties that warrant the equations that they set up, or their expectations about whether and how other U.S. high school geometry teachers will ask their students to provide such justification, dependent on whether the problem has a geometrically-meaningful solution?
   b. Are experienced geometry teachers' attitudes towards students' decisions to write down the geometric properties that warrant the equations that they set up, or other U.S. high school geometry teachers' decisions to ask them to do so, dependent on whether the problem has a geometrically-meaningful solution?

As explained in chapter 3, the basic design of the multimedia questionnaires, including the types of items that they contain, were modeled after multimedia questionnaires previously designed to investigate other norms, as substantial work had been done to improve and assess their reliability and validity (e.g., Herbst, Aaron, Dimmel, & Erickson, 2013; Shultz et al., 2018). However, the overall design was novel, as previous instruments were each designed to investigate a single norm. In order to investigate the relationship between two norms, I needed to adapt the methodology. I did so by including two segments in each storyboard—one in which the GCA-Figure norm was either followed or breached, and one in which the GCA-Theorem norm was either followed or breached—and questions about each segment of the storyboard in each questionnaire. To vet these adaptations, in 2015, I conducted a pilot study in which I administered this questionnaire to a sample of 40 secondary school mathematics teachers from a midwestern state (Boileau & Herbst, 2015; 2016). That study provided some preliminary evidence that the GCA-Theorem norm exists and reason to believe that U.S. high school geometry teachers deem decisions to follow it more acceptable than decisions to breach it. It failed to provide evidence that the GCA-Figure norm exists, but revealed issues with the design of the questionnaires that explained why breaches of that norm were not remarked, which I used to revise the instrument before administering it to the national sample.

With that explained, in the next chapter, I describe the earlier research on norms of mathematics instruction alluded to throughout this chapter as well as the literature on geometric calculation. In chapter 3, I elaborate on the brief description of the present study just presented.

Chapter 2 - Literature Review

In the previous chapter, I first framed the present study as a study of norms of mathematics instruction then specified that it is a study of some of the norms of an instructional situation in U.S. high school geometry called geometric calculation with algebra (GCA). I also noted that GCA is one of two common types of geometric calculation (GC)—the other being geometric calculation with number (GCN)—and that GC, GCA, and GCN are terms used in the literature to refer to not only instructional situations, but also mathematics problems and mathematical tasks (i.e., work on mathematical problems)[18]. These distinctions are relevant to the present study as they are useful in describing the current state of the literature on GCA. Currently, there are no other studies *of* GCA (other than the pilot of the present study; Boileau & Herbst, 2015, 2016)—no studies of the instructional situation of GCA, GCA tasks, or GCA problems—let alone studies of their norms. There are, however, a few studies of student thinking that use GCA problems to investigate it (Dindyal, 2007; Suwito et al. 2016) and curriculum studies that evidence the presence of GCA problems in certain curricula (Chang, 2013; Stacey and MacGregor, 1999). Moreover, there are a handful of studies of GCN tasks (e.g., Aleven et al., 1998; Küchemann & Hoyles, 2002; Hsu, 2010a; Hsu & Silver, 2014), as well as studies of student thinking that used GCN problems (Ayres & Sweller, 1990; Lawson & Chinnappan,

[18] This definition of task comes from Herbst (2006), who distinguished between the concepts of *problem*, *task*, and *instructional situation*. Specifically, following Brousseau, he defined a *mathematical problem* as "a question whose answer hinges on bringing to bear a mathematical theory within which a concept, formula, or method involved in answering the question is warranted" (p.315). Following Doyle, he defined a *task* as "the universe of possible operations that an individual might or might not take while working on a problem, toward a certain product, with certain resources, the feedback that the problem can provide on those operations, and the operations adapted to the feedback that may ensue" (ibid., p.315). Finally, as mentioned in the introduction, he defined an *instructional situation* as "any one of the customary ways in which classroom actions are framed into units of work so as to be traded in for (or accounted to) claims over the knowledge at stake" (ibid., p.316).

2000) and curriculum studies that evidence their presence in certain curricula (e.g., Hunte, 2018; Lang & Ruane, 1981), which are relevant to justifying the design of the present study and contextualizing its results. After reviewing that literature, I review relevant literature on norms—focusing mainly on norms of mathematics instruction, but acknowledging relevant literature from other fields—in order to demonstrate what one might gain from considering mathematics instruction as an activity regulated by norms as well as how someone convinced of this (such as myself) could approach the study of instructional norms.

2.1. A review of the literature on GC

While there are very few references to GC—in particular, to GCA—in the literature[19], they have provided some useful insights. One that might be taken for granted, but is useful, is the evidence that GCA tasks require knowledge of both geometry and algebra. This may be taken for granted because this characteristic of GCA tasks is now used to define them and differentiate them from other types of mathematical tasks (e.g., Herbst, 2010; Hsu, 2010a). Arguably the best evidence that GCA tasks require knowledge of both geometry and algebra comes from Dindyal (2007), who had six U.S. high school geometry students' work on two GCA problems in the context of a set of task-based interviews. He claimed that the tasks required "the use of variables and unknowns, the writing and solution of simple linear equations, the writing and solution of linear simultaneous equations in two unknowns, the substitution of values in expressions, and the recall and use of formulae within geometry" (p.77). His evidence was that participants with more geometric knowledge were better at remembering relevant geometric formulae, students with more algebraic skill were better at solving equations, and students that had both were most successful, while those who lacked both struggled. Participants in this study were selected so that

[19] By that name or any other.

they would differ in terms of their knowledge of algebra and geometry, based on their performance on algebra and geometry pre-tests.

Additional—albeit less-compelling—evidence that GCA problems require knowledge of both geometry and algebra was provided by Suwito et al. (2016). In this study, two 9th grade students in Indonesia (similar in age to the participants in Dindyal's study, as most students in the U.S. take geometry in 9th grade or 10th grade) completed two problems that the authors described as an example of "algebra questions which [are] mixed with geometry" and an example of "geometry problems [that] can be solved algebraically" (p.25). This work was also produced in the context of a set of task-based interviews. The problems were chosen from a larger set that included problems from the PISA (2009) assessment and problems designed by one of the researchers. One problem presented students with three stacks of shapes: rectangles and hexagons. The height of two of the stacks were given and the task was to determine the height of the third stack, which could be accomplished by labeling the height of each shape as a variable and setting up a system of equations, as both students did. I would argue that this problem is not a GCA problem, however, because it does not require the students to use any knowledge of geometry. The problem could be solved—and was solved by both students—by defining one variable to represent the height of each hexagon and another variable to represent the height of each rectangle, writing an algebraic equation that represents the height of each of the first two stacks as a linear combination of those variables, solving that system of equations to determine the values of those variables, writing an algebraic expression to represent the height of the third stack, then using the solutions to the system of equations to evaluate that expression and thereby determine the height of the third stack. I say this did not rely on knowledge of geometry because the same procedure could have been use regardless of the type of shape in the stacks.

The other problem was a word problem that described the relationship between the dimensions and areas of a square plot of land that contained a rectangular garden. It could be solved—and was solved by both students—by modeling the plot of land and garden as squares and rectangles, respectively representing the known length of the larger square as a variable and the dimensions of the garden as a function of x, then solving that quadratic equation. The work of solving this problem fits Hsu's (2010a) definition of a GCA task in that it involved "calculations done within mental... geometric diagrams on the basis of geometric principles or formulae" (p.16) and required "algebraic skills... to calculate the unknowns" (p.18). I claim that this is less-compelling evidence that GCA problems require knowledge of geometry and algebra than Dindyal (2007) provided, because Suwito et al. did not demonstrate that the problem could not be solved without such knowledge. However, they did show that such knowledge is sufficient: While the work that the two students produced in their attempts to solve the second problem is purely algebraic and variables were not defined, each student's interpretation of the solutions to their equation in their final answer implies that the variable in their equation represents the length of the sides of the square plot of land.

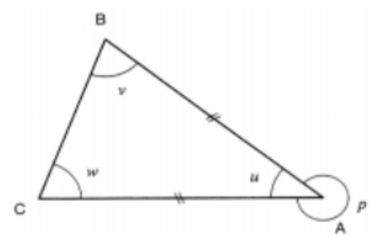
As mentioned in the introduction chapter, Hsu (2010a) demonstrated that Taiwanese students' performance on GCN tasks was associated with their performance on geometric proof (GP) tasks, with which students often struggle (Reid & Knipping, 2010). She then used that evidence to argue that "Taiwanese students' prior experiences in solving GCN tasks have a significant impact on their competence in creating GP" (p.214), given that both types of tasks require a common set of knowledge and skills (e.g., deductive reasoning with diagrams). While one might conjecture that work on GCA tasks could be similarly useful at improving students' abilities to compose geometric proofs, she suggested that "the extra algebra work necessary in solving GCA tasks can increase the cognitive demand of the tasks and confound the conceptualization of the relations between GC and GP, especially when algebra work is a difficult learning topic for students” (Hsu, 2010a, p.16). While this has yet to be empirically verified, even if it were true, the literature on GCA also reminds us that the algebraic work required to solve GCA problems could be allocated to technology. In particular, Schumann and Green (2000) describe four ways GCA problems can be solved using Dynamic Geometry Environments (DGE) and Computer Algebra Systems (CAS).

Next, the literature suggests that, while GCA problems can be designed to require students to set up and solve systems of linear equations and quadratic equations (Dindyal, 2007; Herbst, 2010; Stacey and MacGregor, 1999; Suwito et al., 2016), it is more common (at least) in some countries for them to only require students to set up and solve a single linear equation in one variable. For example, based on their analysis of mathematics textbooks used in Australia, Stacey and MacGregor (1999) concluded that typical GCA problems “require no knowledge of algebra at all beyond the solving of easy equations (e.g., $2x+3x=120$ and $3x=150$, substantially below Year 10 level)” (p.66). This is consistent with my perusal of U.S. high school geometry textbooks (e.g., Cummins et al., 2003) and observation of U.S. high school geometry classrooms. It is relevant to the present study as the GCA problems that I designed for the present study, which I describe in the next chapter, are like this.

As mentioned earlier, the literature on GCN also provides some relevant insights into GCA tasks and problems. For example, there is some evidence that students are not expected to provide the geometric reasoning behind their calculations when solving GCN problems. In particular, Küchemann and Hoyles (2002) describe students’ responses to a particular GCN task (Figure 2.2) that was included in a “written proof test” (p.43), administered to a large random sample of year 8 and year 9 English secondary school students to the GCN problem, in the context of the Longitudinal Proof Project.

The diagram shows. Triangle ABC.

Side AB is the same length as side AC.



A) Find the size of angle v, when angle p is 320°.

Write down each step of your calculations.

B) Write down your first step again and give a reason for the step.

C) Write down your next steps again and give a reason for each.

Figure 2.1: Diagram and questions from Küchemann and Hoyles (2002, p.45)

As shown in Figure 2.2, the problem included both a request to calculate one of the interior angles of an isosceles triangle given one of its exterior angles as well as a request to justify the equation they set up and a request to justify the steps taken to solve the equation. However, they found that, “students often gave reasons which were procedural… rather than, or as well as, giving a geometrical justification for their step” (p.46). I take this as evidence that students may not be expected to provide the geometric reasons behind their calculations in school, as that is one possible explanation for why students would not have provided such justification, even when asked to provide justification by this task. This is relevant to the present study as it was one of the reasons I conjectured that students—at least in the instructional situation of GCA in U.S. high school geometry—are not expected to document their geometric reasoning when completing a GCA task either (the other reason being my perusal of GCA problems in U.S. high school geometry textbooks and observation of U.S. high school geometry classrooms).

While it may be seen as disappointing that students may not be expected to document the geometric reasoning behind their calculations when solving GC tasks (e.g., because their requirement to use such reasoning may be seen as the only thing that warrants their inclusion in a geometry curriculum), there is some evidence that there is room for a teacher to breach that norm and even GC problems can be used to teach new geometric ideas (rather than to assess whether students know a previously taught property and can use it to solve a problem). This evidence was produced by González and Herbst (2013), who describe a classroom scenario in which students were solving a GCA problem that required knowledge of a theorem that had not yet been taught, the teacher asked a student at the board to justify an equation in their solution, a student conjectured that theorem, then the teacher reframed the task as one of proving that conjecture. This study also demonstrates the usefulness of Systemic Functional Linguistics (Halliday, 1978; Halliday and Matthiessen, 2004) in understanding mathematics instruction, which was one of the reasons that I use this theory to analyze the open-response data (i.e., responses to the open-ended items in the questionnaires) that I collected.

Yet arguably the most relevant—and, in my opinion, rigorous—literature on GCN is Hsu's (2010a) dissertation and the set of associated papers (Hsu, 2007, 2010b; Hsu & Silver, 2014). Her dissertation reported on three studies. The first was a study of the GCN tasks included in one Taiwanese mathematics teacher's curricular/instructional materials. The second was a study of students' completion of GCN tasks in one Taiwanese mathematics teacher's class and of the ways the teacher managed that work. Hsu (2010a) framed these two studies as investigations of what Stein, Grover, and Henningsen (1996) describe as the stages of task implementation: tasks in "curricular or instructional materials", tasks "set up by the teacher in the classroom", and tasks "implemented by students during the lesson" (p. 459).

The third of Hsu's studies was an investigation in which 8th- and 9th-grade Taiwanese students were asked to complete a set of GCN and geometric proof (GP) tasks that had the same level of cognitive demand[20], that included diagrams that had the same level of complexity[21], and that required knowledge of the same geometric properties. As alluded to earlier, she conducted this study in order to determine whether work on GCN tasks could develop students' capacity to construct geometric proofs. From study 3, she found that students' performance on GCN tasks was predictive of their performance on GP tasks, which she attributed to both types of tasks requiring similar knowledge and skills. Based on the results of the studies, she argued that providing students with more opportunities to complete GCN tasks could improve students' performance on GP tasks:

> Combining students' performances of both grade levels on GCN and GP seems to suggest a learning trajectory to facilitate students' proficiency in creating GP because of the diagram complexity and problem-solving complexity GCN tasks used and enacted in Taiwanese classroom (see Chapter Two and Three). When solving GCN tasks, the complexity of diagram provides students opportunities to learn strategic knowledge, which is a key for effective problem solvers who need to recall actions that are likely to be useful when choosing which actions to apply among several alternatives (Weber, 2002; 2005). Hence, class experiences in solving GCN tasks prior to formally learning GP can nurture students' strategic knowledge in terms of visualizing diagram configurations so that they can retrieve relevant geometric properties and combine

[20] Stein, Grover, & Henningsen (1996) define cognitive demand of a task as the complexity of the thinking required to complete it. They propose three level: "memorization", "use of procedures and algorithms (with or without attention to concepts or understanding)", "employment of complex thinking and reasoning strategies that would be typical of 'doing mathematics' (e.g., conjecturing, justifying, interpreting, etc.)" (p.461)

[21] Hsu's (2010a) definition of diagram complexity is quite complicated. As it is not used in the present study, readers interested in task analysis are recommended to consult p.37 and p.49-51 of that text.

> different geometric properties to generate a valid solution strategy in novel ways. This training with working on GCN tasks with diagram complexity can later contribute to the learning of GP.
>
> (p.197)

In the introduction, I argued that, since GCA tasks require similar geometric reasoning, they may also prove useful in supporting students in developing some of the knowledge and skills that they need to compose geometric proofs.

As alluded to above, Hsu wrote two conference papers (Hsu, 2007, 2010b) and one journal article (Hsu & Silver, 2014) based on that work. Hsu (2007) reports on her analysis of a segment of one lesson in which the teacher assigned their students a GCN task that requires *transformational observations*—"operations on objects and anticipations of the operations' results" (p.258)—similar to geometric proofs. Hsu (2010b) is a short version of the third paper in her dissertation. Finally, Hsu and Silver (2014) analyzed the sources of the GCN tasks used by the teacher whose class Hsu observed as well as "two dimensions of the complexity [of those tasks]: diagram complexity and problem solving complexity" (p.464). They found that the tasks she used were from a textbook, a workbook, and supplemental materials that she created, and that the problems she created tended to be more complex than the ones in the textbooks.

This concludes the review of the literature on geometric calculation. As suggested earlier, one of the main take-aways from it is that this literature is thin. However, from it, we see that most research on GC has focused their prominence in various curricula as well as students' performance on GC tasks outside of their regular classrooms. An exception is the second study reported in Hsu's dissertation, which included analysis of the teacher's management of the work on GC (specifically, GCN) tasks. In that sense, that study is the most similar to the present study—a study of mathematics teaching practice. Yet it differs in that it was based on analysis of classroom video rather than survey data. It also differs in that its focus was on how GCN tasks may be adapted throughout the stages of task implementation, as opposed to on the norms that regulate the GCA tasks that teachers assign, how students solve them, and how teachers' respond to proposed solutions. For the latter reason, in the next section, I review literature on norms, with a particular focus on norms of mathematics teaching.

2.2. A review of the literature on norms

Norms have been a topic of research in several fields—such as sociology (Gibbs, 1965), philosophy (Bicchieri & Muldoon, 2014), and social psychology (Cialdini & Trost, 1998)—for several decades. As a consequence, the literature on norms is currently very large—sufficiently large, in fact, to have warranted a synthesis of the existing literature reviews (most of which covered the literature from only one or two fields). In conducting that review, Legros and Cislaghi (2020) found that there is considerable variation, yet also noteworthy commonalities, in (1) how norms are defined, (2) how they are believed to influence behaviour, (3) how they are believed to evolve (including how they come into existence and dissolve), and (4) "whose behavior and (dis)approval [are believed to] matter in sustaining the norm" (p.75). As the goal of the present study is to uncover norms and gauge attitudes towards norm-compliant and norm-defiant behaviour, the first of those four themes (defining norms) is arguably the most relevant to elaborate here.

In terms of how norms have been defined, Legros and Cislaghi (2020) explain that researchers tend to agree that norms are implicit rules that are "shared by some members of a group", can be either "[p]rescriptive or proscriptive", are "[r]elated to behaviors and inform decision making", and are "[c]apable of affecting the health and well-being of groups of people"

(p. 65). Accordingly, researchers tend to agree that they are *not* “[i]nstinctual or biological reactions”, “[p]ersonal tastes”, “[p]ersonal habits”, or “[b]ehavioral regularities in a group due to demographic trends, common choices made under very limited options, or the aggregation of individuals with similar tastes” (ibid.). That said, Legros and Cislaghi (2020) also explain that there remains some disagreement. In particular, there is some disagreement amongst researchers about how (i.e., in what sense) norms are shared by members of a group: While some researchers have suggested that they are shared in the sense that they “stem from human interactions”, others suggest that they are shared in the sense that they are “expectations about other people’s beliefs and behaviors”, or that they “hold social meaning”, or that they “allow the functioning of the social structure” (p.65).

A shortcoming of this important review, however, is that it does not describe the variety of research methods that are used to investigate norms (e.g., what the norms of a given situation are, whether the people to whom they apply tend to approve or disapprove of decisions to follow or breach them). In fact, some of the earlier reviews do not either (e.g., Gibbs, 1965). For that reason, research methods will be a focus of the present review. However, as the entire literature on norms is too vast to cover within the scope of the literature review chapter of a dissertation, I focus here on the research on norms within mathematics instruction.

While norms have been a much less popular topic of research in mathematics education than in other fields (such as, sociology, philosophy, and social psychology), there have been a number of significant contributors (as well as contributions). For that research, the remainder of this literature review will outline each of their programs of research, noting important similarities and differences along the way.

2.2.1. The Didactical Contract

One of the earliest and most influential contributions to the literature on norms of mathematics instruction was Guy Brousseau's *Theory of Didactical Situations* (TDS); in particular, his notion that, when students are engaged in a mathematical activity by their teacher, both parties' actions (and interactions) are bound by an implicit *didactical contract* (Brousseau, 1997; Brousseau, Brousseau, & Warfield, 2014; Brousseau & Warfield, 1999). While Brousseau does not use the word *norm*, the didactical contract stipulates what each party (teacher and student) can expect from the other and can thus be described as a set of norms (e.g., Herbst, 2003). The importance of the notion of the didactical contract (like the more-general notion that social behaviour is regulated by norms) is that it reminds us that individuals are not completely agentic, that their decisions are influenced—both constrained and made possible—by the contexts[22] within which they are made[23]. As a consequence, Brousseau argued that students—and I would add teachers—should not be criticized as if this were not the case; that is, as if they could do whatever they wanted without consideration of social accountability.

The main empirical evidence of the existence of the didactical contract that Brousseau produced was "a small number of clinical, didactical tutorial sessions" (Brousseau & Warfield, 1999, p.3) conducted between 1976 and 1983 with a group of nine students. These students were chosen for the sessions because they were attending the public lab school that he helped create (the COREM) and were failing their mathematics class. This work is arguably best represented by his seminal paper, The Case of Gaël (Brousseau, 1981, 2009; Brousseau & Warfield, 1999). In that paper, Brousseau demonstrates how one student's lack of engagement, particularly his unwillingness to justify his work or suggestions that he was unable to do so, was due to his

---

[22] I.e., *environments* (Cohen, Raudenbush, & Ball, 2003)
[23] And anticipations of those contexts, when decisions are planned in advance.

desire to avoid confrontation and understanding of the situation as one in which he could escape confrontation by deferring to the classroom's custodian of mathematical knowledge—the teacher. The evidence provided is quite compelling: The researchers demonstrate that, through "introduc[ing] more and more difficulties and put[ting] [the student] under the obligation of surmounting them", the tutor was able to have the student—Gaël—both take responsibility for solving problems and, thus, construct (i.e., learn) the target mathematical knowledge (Brousseau & Warfield, 1999, p. 45). In their first session, the tutor asked him to attempt to solve the following problem from his workbook, which he (and his classmates) had previously been assigned by their regular teacher: "In a parking lot there are 57 cars. 24 of the cars are red. Find the number of cars in the parking lot which are not red." (ibid., p.10). Gaël completed the task in the same way that he had in his textbook, 57 + 24 = 81, explaining that he was "going to do what I learned from the teacher", despite receiving feedback from his teacher that this solution was incorrect. Through subsequent interactions with Gaël, the tutor is able to get him to create several representations of the problem and, from them, is able to determine (1) that Gaël had not properly understood the problem—"that he ha[s] difficulty in envisaging that there is only one set of cars, with two properties: being in the parking lot and being red" (ibid., p.11)—and that he lacked "the ability to use subtraction to find the two terms of the addition" (ibid., p.12).

Next, to rule out the possibility that the issue was larger—that Gaël was unable to simultaneously consider entities and their parts—the tutor provided him with some wooden beads, some of which were red and some of which were green, and asked him "whether there are more wooden beads or red beads and to justify his answer" (ibid., p.14). Gaël answered that there were more wooden beads (his justification is not reported). Then, to confirm that Gaël understood that addition is commutative (as they imagined he did), the tutor gave Gaël three rods—labeled *A*, *B*, and *C*. Rod *B* was placed to the right of rod *A*, end to end, and rod *C* was placed just above them, demonstrating that the combined length of rods *A* and *B* equaled the length of rod *C*. Then the tutor removed rods *A* and *B*, aligned the left side of rod *B* with the left side of rod *C* and asked Gaël what would happen if they placed rod *A* at the end of rod *B*. He answered correctly: the right side of rod *A* would align with the right side of rod *C*. However, to investigate his willingness to debate his answer, the tutor then told him that another student said that the combined length of *B* and *A* would actually be less than the length of *C*. He was unwilling to debate, claiming that he could not say who between them was correct. Brousseau offers the following interpretation:

> This type of conduct on this test characterizes children whom Gréco situates in an intermediate (pre-operational) stage, where the structures of the subject are still under construction, so that the compensations are incomplete and fragile. This cannot be the case with Gaël. It would seem that without any doubt, his sudden absence of conviction in the face of a simple counter-proposition has to do with his general attitude towards others around him when his own knowledge is put into play.
> (ibid., p.15)

The paper then includes descriptions of other sessions in which Gaël engages in other mathematical work strategically chosen by the tutor to continue to (1) determine the current state of and/or develop Gaël's knowledge, (2) demonstrate that it was the didactical contract of his regular classroom that caused his aversion to discussions around his mathematical work, and (3) have Gaël realize that "knowledge was not to be found in either the discourse or the desire of the teacher, but rather in a relationship with the milieu[24]" (p.45).

As mentioned earlier, one reason I offer this detailed description of the Case of Gaël is that it exemplifies the type of empirical evidence that Brousseau produced to support his claim that students and teachers act as if bound by an implicit didactical contract. However, a second reason is that I imagine it will help some readers see the connection between this line of research and the one that I review next—the set of constructivist teaching experiments that led to the development of the idea of sociomathematical norms and general classroom social norms.

2.2.2. General Classroom Social Norms and Sociomathematical Norms

Cobb, Yackel, and Wood's group is one of the best-known contributors to the literature on norms of mathematics instruction. In terms of theory, one of their main contributions to this literature has been their distinction between what they call *general classroom social norms* (Cobb, Yackel, & Wood, 1989; Yackel, Cobb, & Wood, 1991) and *sociomathematical norms* (Voigt, 1995; Yackel & Cobb, 1996). The former concept was developed by Yackel, Cobb, and Wood in the context of a constructivist teaching experiment conducted in a second-grade classroom in the U.S.. The latter was developed by Voigt in a similar context. In both cases, norms were inferred from observed tensions that arose as a teacher attempted to engage students in a student-centered form of instruction intended to increase their sense of autonomy. In particular, the researchers noticed that the teacher needed to negotiate expectations about the students' responsibilities, given that the students had developed their expectations in traditional classrooms—where the teacher was the custodian of the knowledge. Yackel and Cobb (1996)

[24] An everyday word in French to which Brousseau adds the following technical meaning to specify his theory: "everything [within a situation] that acts on the student or that she acts on" (Brousseau, 1997, p.9). Examples include the mathematical problem that the student is working on, the resources that they may use (e.g., wooden beads) and, in *didactical situations* (as opposed to *adidactical situations*), the teacher.

distinguished between general classroom social norms and sociomathematical norms by explaining that, while a general classroom social norm might apply to multiple courses of study (e.g., mathematics and social studies), sociomathematical norms are specific to mathematics. As examples of general classroom social norms, they list, “[t]he understanding that students are expected to explain their solutions and their ways of thinking” and “the understanding that when discussing a problem students should offer solutions different from those already contributed" (ibid., p.461). In contrast, they explain that sociomathematical norms would include, “normative understandings of what counts as mathematically different, mathematically sophisticated, mathematically efficient, and mathematically elegant in a classroom”, as well as “what counts as an acceptable mathematical explanation or justification” (ibid, p.461).

While the group’s collaborative efforts to investigate norms of mathematics classrooms subsided by the late 1990s, Cobb remained interested and later tied this work to the research of mathematics identity that was gaining popularity (e.g., Gee, 2001; Martin, 2000). From my perspective, this new line of work is a natural—and important—extension of his earlier work, as well as Brousseau’s: Rather than providing students with opportunities to develop deep and personal understandings of mathematics, most classrooms leave students with the impression that the teacher is the authority on mathematics and that the student’s work is limited to learning and correctly implementing a set of algorithms. For example, Cobb, Gresalfi, & Hodge (2008) make this claim when describing their analysis of a design experiment that they conducted in a high school algebra classroom:

> [T]he students' primary obligation as they understood it in the algebra class was to produce correct answers by enacting prescribed methods on written notations. The

> students therefore viewed themselves as limited to exercising disciplinary agency. Their interview responses also indicate that, from their viewpoint, authority was distributed primarily to the teacher, who was solely responsible for determining the legitimacy of their solutions. Most of the students viewed themselves and most of their peers as not being able to fulfill the teacher's expectations, primarily because the criteria for what counted as an acceptable solution appeared to be arbitrary from their perspective. (p.62).

Next, I turn to Stigler and Hiebert's work, which contrasts interestingly with Cobb, Yackel, and Wood's, as well as Brousseau's, in terms of the approach that they took to studying norms. Namely, rather than use teaching experiments to elicit students' expectations, they inferred teachers' expectations from trends that they observed in the practice of large samples of teachers.

2.2.3. Cultural Scripts for Teaching

Other important contributions to the literature on norms of mathematics instruction were made by Stigler, Hiebert, and colleagues in the context of the Third International Mathematics and Science Study (TIMSS). As mentioned in the introduction, through analyzing large sets of classroom video from the U.S., Germany, and Japan, they noticed strong patterns of behaviour amongst teachers and students from the same country and clear differences between the behaviour of teachers and students from different countries. For example, they found that "U.S. mathematics teachers tend to teach students how to solve a particular type of problem, then ask them to solve examples on their own", while "Japanese teachers often give students problems to work on that they have not seen before" (Stigler, Gallimore, & Hiebert, 2000, p.88). Their general explanation for this was shared expectations amongst teachers and students from each country. At times, they would refer to these as norms. For example, Stigler & Hiebert (1997) wrote:

> Teaching is a cultural activity. It is an everyday event that occurs throughout all parts of American society. Over time, we have developed *norms* and expectations for teaching that are widely shared and passed along as one generation of students becomes the next generation of teachers. Because our models of how teaching should look are so widely shared and so familiar, they become nearly invisible. We come to believe that this is the way teaching must be. When we observe teaching in other countries, these accepted and unquestioned practices are brought to light…
> (p.14, *my emphasis*).

However, they more commonly attributed these patterns in behaviour to *cultural scripts for teaching*. They defined these as "generalized knowledge about [an] event that resides in the heads of participants" that "not only guide behavior", but "also tell participants what to expect" (Stigler & Hiebert, 1998, p.4). That said, the two ideas are highly related; for example, Cristina Bicchieri, a leading scholar on social norms, describes norms as embedded in scripts (Bicchieri, 2006). But, more importantly, whichever term Stigler and Hiebert use in a given paper, their general claim is the same: The patterns of behaviour that they observed are due to shared expectations.

Before moving on to the next section, it is worth noting two important differences between Stigler & Hiebert's research on norms (and scripts) and the research on norms conducted by Brousseau, as well as Cobb, Yackel, and Wood. First, as mentioned earlier, Stigler & Hiebert's research differs from both Brousseau's and Cobb, Yackel, and Wood's research in that they inferred expectations from observed trends in behaviour rather than from reactions to (intentional and unintentional) breaches of norms. Since norms are rules that people expect each other to follow and follow for that reason, rather than "[b]ehavioral regularities in a group due to demographic trends, common choices made under very limited options, or the aggregation of individuals with similar tastes" (Bicchieri, 2006; Legros & Cislaghi, 2020, p.65), evidence that individuals' expectations are not met is better evidence of a norm than a pattern of behaviour. That said, an advantage of Stigler and Hieber's work is that they analyzed large, random samples of lessons, which enabled them to generalize their findings beyond their sample. Specifically, their sample design is one of the main reasons they were able to make claims about how mathematics is taught in various countries. In contrast, the tutoring sessions that Brousseau studied, and the classrooms Cobb, Yackel, and Wood's studied, were not only few in number, but also quite unlike regular mathematics classrooms. That said, they served their purpose: to develop theory (which could later be tested with larger samples). In the next section, I describe a final line of research that combined both approaches—one that began with work with small samples of teachers to develop initial versions of a norm-based theory of mathematics instruction and research methodology to test it, which led to work with large-scale samples of teachers in order to validate them.

2.2.4. Norms of Instructional Situations

A fourth major contributor to the literature on norms of mathematics instruction have been Herbst, Chazan, and their colleagues. In terms of theory, their major contribution has been their account of the practical rationality of mathematics teaching (Herbst & Chazan, 2003, 2012;

Chazan, Herbst, & Clark, 2016): A rational reconstruction of mathematics teaching that assumes the work of teaching mathematics "rest[s] on the base of a common practical rationality, upon which individual practitioners can build their own mathematics teaching against the backdrop of their personal commitments and the demands of the institutional contexts where they work" (Herbst & Chazan, 2003, p.2). These demands include expectations to follow the norms of those contexts, unless a deviation is warranted by their personal commitments and/or another institutional demand (Chazan, Herbst, & Clark, 2016). Particularly relevant to this dissertation is their suggestion that, in between norms that are negotiated by a given teacher and their students (described by Cobb, Yackel, and Wood) and norms that influence mathematics instruction in every classroom in a given country (described by Hiebert and Stigler), one will find norms that apply to recurrent *instructional situations* (Herbst, 2006) in all offerings of a given course, such as the instructional situation of GCA in U.S. high school geometry (Herbst, 2010).

> Even for the same course of studies, say high school geometry, different contracts could... stipulate the roles and responsibilities of teacher and student differently. In particular, it is conceivable that some contracts might include the expectation that every new task would require negotiation about how the general norms of the [standard] contract apply (e.g., What is required of the teacher to get students to work on a particular task? What does it mean for students to work on that task?). It is also conceivable, and we argue more likely, that contracts rely on a manifold of instructional situations that forego the need for some of those negotiations much of the time.
>
> (Herbst & Chazan, 2011, p.422-423)

While one may read these statements as implying that an instructional situation is a recurrent type of mathematical work (typically done by the student) and this is partly true—it would be fair to characterize and differentiate between two instructional situations (e.g., GCA and doing proofs in U.S. high school geometry) on the basis of the work typically produced in each type of situation—the full definition also stipulates how such work would be interpreted by the teacher:

> I use *instructional situation* to refer to any one of the customary ways in which classroom actions are framed into *units of work* so as to be traded in for (or accounted to) claims over the knowledge at stake (and, reciprocally, any one of the customary ways in which the teaching or learning of objects of knowledge is deployed as classroom work).
>
> *Situation* (short for *instructional situation*) identifies each of the frames participants use to know who had to do what and when, so that whatever they do can be used to claim the fulfillment of their contractual obligations.
>
> (Herbst, 2006, p.316)

In other words, an instructional situation is a recurrent type of *instructional exchange* of students' mathematical work on a given familiar type of task (Doyle, 1988) and a teacher's claim on the extent to which the students seem to have acquired the knowledge and skills that the task was assigned to evaluate on the basis of their work (Herbst & Chazan, 2012). Based on this definition, the instructional situation of GCA can be defined as one in which "[t]he work to be done includes finding deductively the numeric value of an unknown (and of expressions that use that unknown) where some of those expressions using the unknown are given as representations of the measure of an element in a figure" is exchanged for "a claim on students' capacity to use a property they already know as well as a claim on maintaining knowledge of algebra skills" (p.53).

In terms of how Herbst, Chazan, and colleagues define norms, while they are of course briefer in some papers than others, I would argue that their conception is best captured by the following general definition of norms and subsequent definition of norms of instructional situations:

> [T]he norms we speak about are implicit regulations of the joint activity of teacher and students, as seen from the perspective of the observer. For us, normative means not necessarily "correct" or "appropriate" from anybody's perspective but rather "unmarked"; a manner of doing joint activity that is ordinary, in the sense that it goes without saying. (Herbst, Chen, Weiss, & González, 2009)

> [N]orms [are] expectations of default behavior that, were they to take place in an instance of an instructional situation, they would go without saying but were they to be breached, they would elicit ad hoc 'repairs'
> (Herbst, Nachlieli, & Chazan, 2011, p.227).

In terms of the types of norms that regulate the actions of teachers and students in these situations, Herbst and Chazan (2012) claim that there are norms of the (standard) didactical contract or *contractual norms* (e.g., that teachers will assign problems, which students are expected to solve, as Brousseau suggested), norms of instructional situations or *situational norms* (just defined), and *task norms*[25]. Herbst, Chen, Weiss, & González (2009) explain that norms of instructional situations fall into three categories: *exchange norms*, which stipulate "what work is done and what its exchange value is", *division of labor norms*, which stipulate "how labor is shared between teacher and student", and *temporal norms*, which stipulate "how that work is organized over time (the sequence and duration of events)" (p.254).

In terms of methodology, most of their research on norms can be characterized as a series of *breaching experiments* (Garfinkel, 1963; Mehan & Wood, 1975)—an umbrella term used to refer to studies in which "persons violate social expectations to demonstrate the existence of underlying rules governing social behavior" (Rawls, 2011, p. 90) as well as mechanisms for restoring normalcy when these rules are broken (i.e., norms are breached). Since its creation by Harold Garfinkel in the 1960s, researchers from a variety of fields (including sociology, social psychology, and education) have employed this methodology to study norms (although not all of them have used the term *breaching experiment* to characterize their approach or the term *norm* to refer to their object of study). These studies have varied in terms of (1) whether the norms were, a priori, either known, hypothesized to exist, or unknown, (2) whether a norm was intentionally breached by the experimenter or whether only naturally-occurring breaches were observed, (3) whether the study was a true experiment (i.e., whether participants were randomly assigned to different experimental conditions; González, 2009), and (4) how participants' reactions to breaches were captured.

An early example in the mathematics education literature was Much and Shweder (1978) who, when observing nursery school and kindergarten classrooms at the University of Chicago,

[25] Defining this term would require introducing more theory and, given that it is not a construct necessary to define the two norms under investigation, here, I share Herbst & Chazan's (2012) example of a task norm: "If a task involves students' reliance on the feedback of a device such as a calculator, the teacher may need to ensure the calculator has batteries and a functional keypad" (p.607)

looked for occasions when the teachers or students call out and/or evaluate "behavior [that] depart[ed] from the ordinary, expectable, and approvable forms associated with a given context" (ibid., p. 22). They took this approach to studying norms as they agreed with Garfinkel and others after him that:

> [I]t is more fruitful to assume that cultural rules are continually tested, employed, clarified, and negotiated in microscopic moments of everyday life. These moments, which we shall refer to as situations of accountability, are easily overlooked, but it is in these often brief verbal moments that the process of cultural control and negotiation can be directly observed.
>
> (ibid., p.20)

Based on their analysis, they proposed that the rules that regulated behaviour in those classrooms could be categorized into five types: "(1) regulations (or laws); (2) conventions (or customs); (3) morals (or ethics); (4) truths (or beliefs); and (5) instructions (techniques, recipes, or 'know-how')" (p.25).

Rather than rely on naturally-occurring breaches, as Much and Shweder (1978) did, Herbst (2003, 2006) co-planned lessons with U.S. high school geometry teachers in order to conduct a series of "design experiment[s] that induced perturbations on the natural variability of teaching through curricular choices present in replacement lessons or units [that] pursued similar goals." (Herbst & Chazan, 2015, p.276). For example, Herbst (2003) co-planned a lesson with the teacher of a "6-week-long summer course in geometry for students who had just finished middle school" (p. 208). In this lesson, the students were assigned a novel task in which they were to compare the area of triangles without using the area formula that they had learned in middle school (·, where *b* is base of the triangle and *h* is the height of the triangle). The goal of the lesson was to have "students realize that the concept of area exceeds the use of a formula to find what areas are." (ibid., p.216) The goal of this teaching experiment—like Cobb, Yackel, & Wood's, as well as Voigt's, and Brousseau's before them—was to determine what negotiations of the didactical contract this would require. As such, it would be fair to categorize these as breaching experiments as all of these researchers, in a broader sense, were studying reactions to breaches of norms.

To avoid intruding on students' opportunities to learn and be able to customize instances of instruction more than they could through co-planning lessons with teachers, Herbst and Chazan subsequently developed the *virtual breaching experiment* (Herbst & Chazan, 2003, 2011, 2015), which confronted teachers with representations of classroom scenarios in which a hypothesized norm has been breached. Those representations of practice included video records of actual classrooms, as well as animations and storyboards in which the students and the teacher are represented by nondescript cartoon characters (Herbst, Chazan, Chen, Chieu, & Weiss, 2011[26]). Some of these studies consisted of showing videos to focus groups of teachers and examining their commentary on them (e.g., Chazan, Sela, & Herbst, 2012; Herbst and Chazan 2003; Herbst, Nachlieli, & Chazan, 2011; Nachlieli, Herbst, & González, 2009; Nachlieli, 2011). Others consisted of packaging storyboards in multimedia questionnaires that included a series of questions about each storyboard (Buchbinder, Chazan, & Capozzoli, 2019; Buchbinder, Chazan, & Fleming, 2015; Dimmel & Herbst, 2017, 2018; Herbst, Aaron, Dimmel, & Erickson, 2013).

[26] In addition to describing the affordance of using such virtual representations, this article also cites literature that provides evidence that teachers project their own students onto the student characters in such representations of practice, and themselves onto the teacher, thereby making such representations useful for investigating teaching practice, in particular, its norms (Herbst, Nachlieli, Chazan, 2011)

Herbst, Dimmel, and Erickson (2016) called this type of questionnaire an Implicit Norm Recognition (INR) instrument. In both cases, the researchers analyzed participants' responses for evidence that they remarked the breach of the norm and for evidence of positive or negative appraisal (Martin & White, 2005) of the breaches. The former was interpreted as evidence that the norm exists given that non-normative behaviour (being unexpected) tends to attract attention (Mehan & Wood, 1975, p.23); the appraisal scores were interpreted as indication of teachers' attitudes towards the depicted breaches.

By administering these and other survey instruments to large samples of teachers, Herbst and Chazan were able to provide evidence of the existence of various norms of instructional situations of U.S. high school geometry (e.g., Dimmel & Herbst, 2017, 2018; Herbst, Aaron, Dimmel, & Erickson, 2013) and high school algebra (Buchbinder, Chazan, & Capozzoli, 2019; Buchbinder, Chazan, & Fleming, 2015). They have also been able to demonstrate that teachers' awareness of many of these norms increases with experience teaching the courses to which they apply (e.g., Herbst, Aaron, Dimmel, & Erickson, 2013). Last, using a measure of teachers' stances towards breaches of a given task norm that consisted of asking participants to rate the relative appropriateness of two problems that differ only in terms of whether the norm of interest is breached or followed, called the diagrammatic register norm (DRN) instrument[27] (Herbst, Dimmel, & Erickson, 2016; Herbst, Kosko, & Dimmel, 2013), Herbst, Chazan, and colleagues were able to determine that a teacher's willingness to assign a non-normative problem may be related to how much of the knowledge needed to teach mathematics (Ball, Thames, & Phelps, 2008; Herbst & Kosko, 2014) they possess (Boileau, Dimmel, & Herbst, 2016).

[27] The norm that was investigated using this instrument is that the premise(s) and conclusion of the propositions that students are asked to prove in U.S. high school geometry are expected to be stated in the diagrammatic register—that is, in terms of characteristics of the diagram provided, rather than in terms of general mathematical concepts.

Herbst and Chazan (2011) later described how the virtual breaching experiment methodology could be further developed by randomly assigning participants to either a set of storyboards in which the target norm is breached or a set of storyboards in which it is followed.

> [O]ur earlier conception of doing experimental research only abided by the notion of experiment as the deliberate reproduction of a phenomenon. But one could also see at least as a possibility that the modern conception of experiment, which emphasizes reproduction of the phenomenon under controlled conditions by way of random assignment of participants to conditions, could be used to confirm that a norm holds: Imagine having two representations of teaching that differed only in that in one of them (the control condition) a hypothesized norm held while in the other (the treatment condition) the hypothesized norm has been breached. Imagine a sample of practitioners who have a comparable degree of socialization in the practice where the norm is supposed to hold. Imagine randomly assigning those participants to one or another representation. Imagine having a way of gauging their satisfaction with the instruction experienced and comparing both groups in regard to that assessment.
> (ibid., p.414-415)

This recommendation was eventually followed by Justin Dimmel, when he designed his dissertation study of two norms of the instructional situation of doing proofs in U.S. high school geometry (Dimmel, 2015; see also Dimmel & Herbst, 2017, 2018, 2020), under Herbst's mentorship. As suggested, Dimmel (2015) randomly assigned his participants to either a *treatment group* that was shown storyboards in which one of his hypothesized norms was breached or a *control group* that was shown storyboards that represented normative work on a proof problem. These storyboards were paired in the sense that, for each storyboard in the breach condition, there was a storyboard in the control condition that was identical except in that the target norm was followed. Similar to prior INR instruments, after viewing each scenario, participants were asked what they saw happening in the scenario (an approximation of researchers noting participants' reactions to breaches in natural settings), to rate the appropriateness of the teacher's actions throughout the scenario, and to explain that rating. They were also asked to rate the appropriateness of the teacher's actions in each of the two segments of the storyboard as well as to explain each of those ratings. Given that normative behaviour—because it is tacitly expected—tends to go unnoticed, while non-normative behaviour—being unexpected—tends to attract attention (Mehan & Wood, 1975, p.23), he interpreted the proportion of participants who remark when each of his hypothesized norms was breached being higher than the proportion of participants who remark when it was followed as evidence that his hypothesized norms exist. He also compared participants' assessments of the teacher's actions across storyboards in the two conditions and concluded that U.S. high school mathematics teachers deem non-normative behaviour as less acceptable than normative behaviour (as Garfinkel, 1963, suggested).

The present study is a natural next step in this line of research. In short, it consists of the administration of an INR instrument that includes questionnaires in which both of the hypothesized norms are followed, questionnaires in which one of them is breached, and questionnaires in which both are breached, in order to determine whether breaching one norm of a given instructional situation may alter either (1) teachers' expectations that other norms of that situation will be followed or (2) their attitude towards breaches of those norms. In the next chapter, I provide a detailed description of that instrument, the sample to which it was administered, and the approach I took to analyzing the data collected.

Chapter 3 Methods

In this chapter, I describe my dissertation study: the instrument that I designed in order to answer my three research questions, the national sample of U.S. high school mathematics teachers to which it was administered, and the approaches I took to analyzing the data collected. As a reminder, those research questions are the following:

1. Do the GCA-Figure and GCA-Theorem norms exist? That is,
    a. Do experienced geometry teachers expect GCA problems to have geometrically-meaningful solutions?
    b. Do experienced geometry teachers expect students to either only share their algebraic work, or to sometimes also verbally state the geometric properties that warrant the equations that they set up, but not to write those properties?
2. Do experienced geometry teachers deem decisions to breach either norm acceptable or at least as acceptable as decisions to follow them? That is,
    a. Do experienced geometry teachers deem other U.S. high school geometry teachers' decisions to assign problems that breach the GCA-Figure norm acceptable or at least as acceptable as decisions to assign problems that follow it?
    b. Do experienced geometry teachers deem students' decisions to breach the GCA-Theorem norm and/or U.S. high school geometry teachers' decisions to ask students to do so acceptable or at least as acceptable as decisions to follow it?

3. Are experienced geometry teachers more likely to expect that the GCA-Theorem norm will be followed, or to disapprove of breaches of the GCA-Theorem norm, when the GCA-Figure norm is followed (than when it is breached)?
   a. Are experienced geometry teachers' expectations about whether and how students will share the geometric properties that warrant the equations that they set up, or their expectations about whether and how other U.S. high school geometry teachers will ask their students to provide such justification, dependent on whether the problem has a geometrically-meaningful solution?
   b. Are experienced geometry teachers' attitudes towards students' decisions to write down the geometric properties that warrant the equations that they set up, or other U.S. high school geometry teachers' decisions to ask them to do so, dependent on whether the problem has a geometrically-meaningful solution?

3.1. The INR-GCA Instrument[28]

The instrument that I designed is called the Implicit Norm Recognition, Geometric Calculations with Algebra (INR-GCA) instrument. It, like Dimmel's INR instrument (described in the previous chapter), consists of sets of multimedia questionnaires, each of which confronts participants with a storyboard representation of a classroom scenario in which each of the two hypothesized norms is either followed or breached (depending on the experimental condition). Each questionnaire also contains a common set of items about the storyboard presented, which ask the participant to describe what they notice about it and to rate the appropriateness of the

[28] The instrument was designed with resources from NSF grant DRL- 0918425 to Patricio Herbst. All opinions ensuing from the use of this instrument are those of the author and do not necessarily represent the views of the foundation.

teacher's actions. Responses are then analyzed in order to confirm the existence of each hypothesized norm and explore participants' attitudes towards breaches of them. As such, the present study is an example of what Dimmel (2015)[29] called a *virtual breaching experiment with control*. Next, I describe the storyboards in more detail.

3.1.1. The Storyboards

The INR-GCA instrument consists of twelve questionnaires. The storyboard in each questionnaire follows the same general plot. In the first segment, a problem is chosen, a student is selected to present their solution at the board, and the teacher circulates while the students work individually on the problem. The algebraic expressions in the diagram are linear[30] and represent side lengths.[31] The participant is shown an over-the-shoulder view of one student's solution, which follows the GCA-Theorem norm in the sense that it does not include a justification for the first equation in the solution. In the second segment, the student who was selected to present their solution at the board writes an equation and begins to solve it by writing a second (slightly simplified) equation, without stating or writing the geometric property that warrants the first equation. This was intended to signal to the participant that the student was not planning to justify their first equation. It is also worth noting that the solution presented at the board is the same solution seen in the first segment (so that the reader would not have to analyze two solutions), although the student who presents at the board is not always the student whose work we see in the first segment.

---

[29] As mentioned in the literature review, Herbst and Chazan (2011) proposed this methodology; Dimmel—one of Herbst's advisees—was the first to implement it and the one who coined the term.

[30] In other words, all of them have the form $ax+b$ (or, equivalently, $b+ax$), where a and b are integers.

[31] Admittedly, this is a threat to the content validity of any claims about the GCA-Figure and GCA-Theorem norms being norms of the instructional situation of GCA, as we do not know if the results would hold for GCA problems in which the algebraic expressions represent the measures of other elements of a figure (e.g., angles or arcs). This is taken up in the discussion section of the dissertation, however, rather than here, to maintain the focus of this document on the coding scheme.

The teacher then interrupts to ask the student for the geometric property that allowed them to set up the first equation. The student verbally states the property (thereby justifying their approach to solving the problem), then finishes solving the problem. Finally, they summarize their work by explaining that they solved the equation and used the solution to determine the measures of the components of the figure that the problem asked them to determine, and the teacher asks the class to evaluate the proposed solution. In the third segment, which consists of a single storyboard frame, the teacher proposes to move on to another problem, signaling that the discussion of the original problem is complete.

The storyboards also differ in ways that can be used to further characterize them. For one, each storyboard followed one of four storylines. These four storylines differ in terms of the problem that the class works on (e.g., the type of geometric figure represented by the pictorial components of the diagram), how that problem is chosen (e.g., whether a student asks to go over a problem from last night's homework or whether the teacher selects a problem from a worksheet), and how the student who solves it at the board is selected (e.g., whether they are selected by the teacher or they volunteer). The storylines also differ in terms of whether the problem is on the board in the first frame of the storyboard or whether a student writes it while the teacher walks around the room, whether the teacher asks the student to summarize how they determined the solution to the problem after they finish solving it, and how the teacher asked the class to evaluate the proposed solution (e.g., whether students are asked to decide whether the answer is right or wrong and why, whether a specific student is asked if they got the same answer or about how their solution differed from the one presented). The storylines are named after the type of figure represented by the pictorial components of the diagram in the GCA problem being solved: the similar-triangles storyline, the parallelogram storyline, the trapezoid storyline, and the isosceles-triangle storyline. Having multiple storyboards in each experimental condition is important to the content validity (Furr & Bacharrach, 2013) of any claims about the instructional situation of GCA made on the basis of the results, as it includes all of the variation represented by the differences between the storylines.[32] The features of each storyline listed above are also important because the attribution of a participant's remark that a hypothesized norm was breached to their expectation that it would be followed depends on the presence (or absence) of other things that they could have otherwise remarked upon.[33] In the results chapter, I note some of the things that participants noted other than the moments when each hypothesized norm was followed or breached.

The three storyboards that follow each storyline can be further distinguished by the fact that each belongs to a different experimental condition. The three experimental conditions differ in terms of which (if either) of the hypothesized norms is breached. Specifically, in the first experimental condition—hereafter, the compliance-compliance condition—both norms are followed. In the second condition—hereafter, the compliance-breach condition—the GCA-Figure norm is followed and the GCA-Theorem norm is breached. In a third experimental condition—hereafter, the breach-breach condition—both hypothesized norms are breached.

Evidence that the GCA-Figure norm is breached in the breach-breach storyboards is that the solution to the equation set up by the student either results in either (1) two of the side lengths being negative (in the case of the Similar-Triangles and Parallelogram storyboards), (2) one of the side lengths being zero (in the case of the Trapezoid storyboards), or (3) the triangle

[32] Conversely, if one only included one storyboard per experimental condition, a threat to the content validity of any claim about the instructional situation that one would make based on the results of the experiment would be that there is no evidence that the results would be the same if the particular problem, the way in which the solution was taken up by the class, or some other feature of the scenario were different.

[33] Similarly, that claim also depends on each item being worded in a way that is as inviting of comments on these features as much comments on the decisions to follow or breach the hypothesized norms.

inequality[34] being violated (in the isosceles-triangle storyboards). Similarly, evidence that the GCA-Figure norm is followed in each of the compliance-compliance and compliance-breach storyboards is that the solution to the equation set up by the student results in valid side lengths, in the sense that they are positive (real) numbers and satisfy all of the properties of the type of figure represented by the diagram being satisfied. To help the reader envision this evidence, Figure 3.1 includes the solution that is presented in the first segment of each storyboard, then presented at the board (sometimes by a different student) in the second segment.

The GCA-Theorem norm, in contrast, is followed in all storyboards, but then also breached in some of them—specifically, in the breach-breach storyboards. As mentioned earlier, it is first followed in each storyboard by the student whose work we see in the first segment not including a justification for the first equation. It is also followed by the student at the board, in the second segment of each storyboard, not initially justifying the first equation in their solution (which is, again, the same solution shown in the first segment). Finally, the GCA-Theorem norm is followed by the student at the board verbally justifying that equation, after being asked to justify it by the teacher[35]. It is worth noting that, if it is true that it is normative within the instructional situation of GCA for a student to verbally state the geometric property that warrants the first equation in their solution, then it would also be normative for the teacher to ask for that justification when it is not provided (i.e., for the teacher to repair the student's breach of that norm). This is because the teacher (as well as, potentially, other students) is expected to uphold

[34] As a reminder, the triangle inequality states that the sum of the lengths of any two sides of a triangle is less than the length of the third side.

[35] More specifically, in the similar-triangles (2900*) storyboards, the teacher asked, "Omega, what geometric property allowed you to set-up your first equation?". In the trapezoid (2901*) storyboards, the teacher asked, "Alpha, what geometric theorem allowed you to set-up your first equation?" In the isosceles-triangle (2902*) storyboards, the teacher asked, "Omega, what geometric property allowed you to set-up the equation $2x+3=5x-12$?". In the parallelogram (2903*) storyboards, the teacher asked, "Beta, how do you know that $5x+6$ is equal to $14-3x$?"

the norms of an instructional situation, the same way as anyone familiar with any type of situation would be (e.g., waiting in a queue; Mann, 1970; Milgram, 1986). This is relevant to the

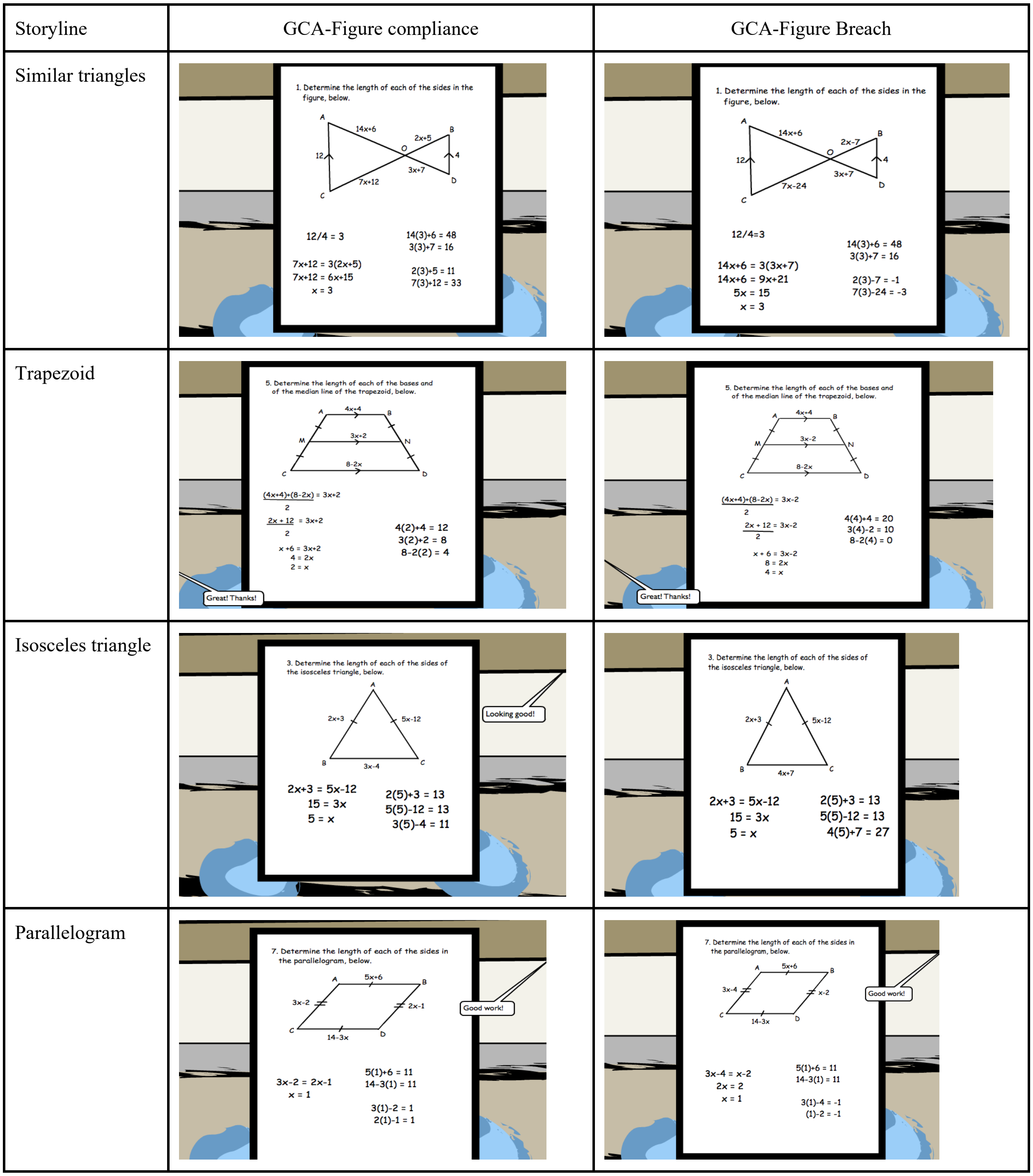
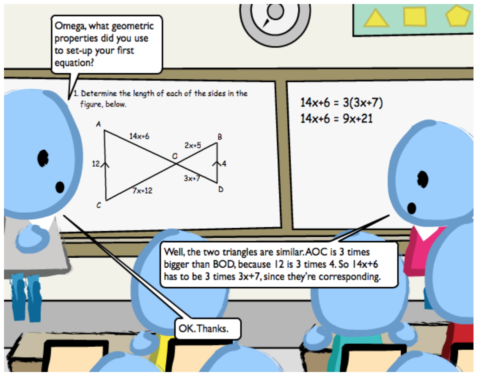
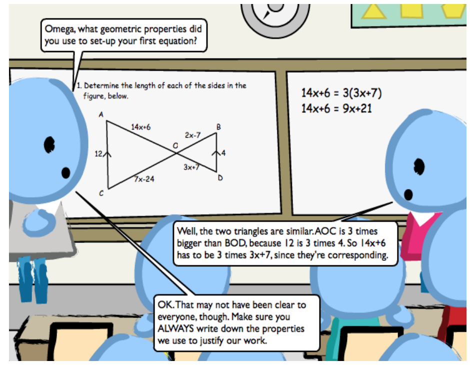


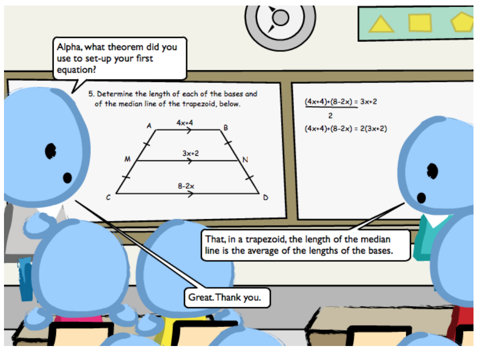
Figure 3.1: Evidence of whether the GCA-Figure norm was followed or breached in each storyboard

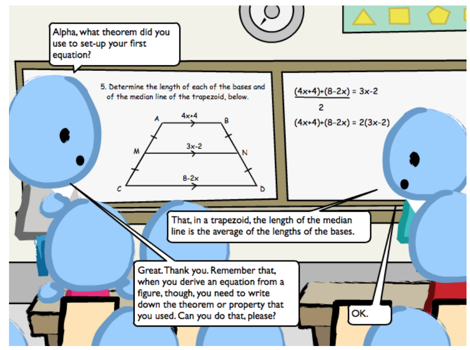
validity of investigating norms that regulate students' solutions by asking teachers to evaluate the teacher's management of work on problems—a point to which I return to in the next section. The GCA-Theorem norm is breached in the compliance-breach and breach-breach storyboards by the student at the board writing the justification of the first equation in their solution, in response to the teacher clarifying that their initial request for justification was actually a reminder for the student to do so. Figure 3.2 resents these interactions between the teacher and the student at the board in each of the storyboards. Again, if it is true that writing the geometric property that warrants the first equation in one's solution is non-normative within the instructional situation of GCA, then so would be the teacher's request for them to do so. Therefore, we would expect participants to remark either the student's decision to write that justification or the teacher's request for them to do so.

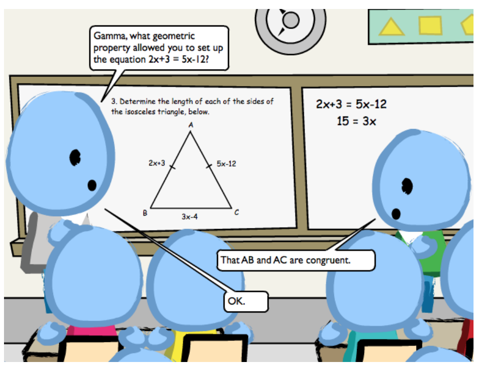
Figure 3.3 represents how each storyboard is the unique combination of one of the four storylines and three experimental conditions. It also includes the names of each of the storyboards in the dataset, which will be used in some of the discussion of the results and important to anyone conducting further analysis of the data. Note here the names of the variables are all 290xy, where x represents the storyline and y represents the experimental condition. More specifically, x=1 indicates that the storyboard follows the similar-triangles storyline, x=2 indicates that the storyboard follows the trapezoid storyline, x=3 indicates that the storyboard follows the isosceles-triangle storyline, and x=4 indicates that the storyboard follows the parallelogram storyline. Likewise, y=1 indicates that the storyboard belongs to the compliance-compliance condition, y=2 indicates that the storyboard belongs to the compliance-breach condition, and y=3 indicates that the storyboard belongs to the breach-breach condition. To provide the reader a full example, storyboard 29032 is included in Appendix A.

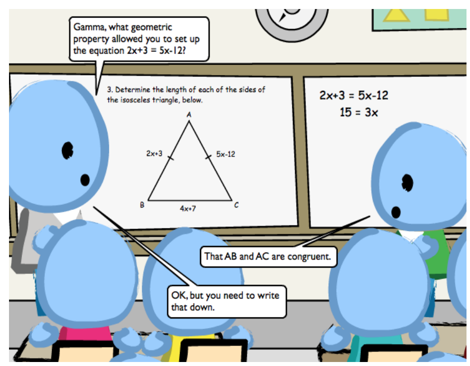
<table>
<tr><th>Storyline</th><th>GCA-Theorem compliance</th><th>GCA-Theorem Breach</th></tr>
<tr><td>Similar triangles</td><td>


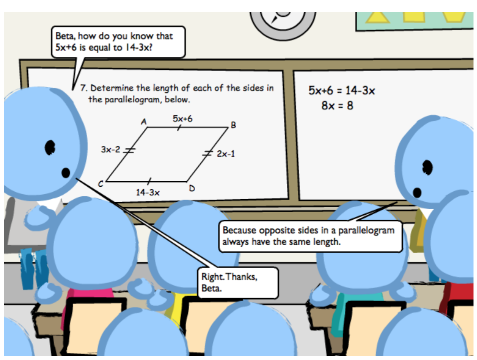
</td><td>


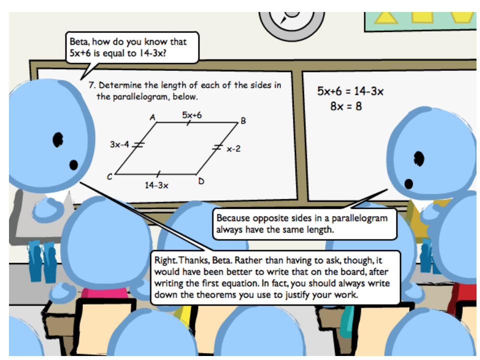
</td></tr>
<tr><td>Trapezoid</td><td>


</td><td>


</td></tr>
<tr><td>Isosceles triangle</td><td>


</td><td>


</td></tr>
<tr><td>Parallelogram</td><td>


</td><td>


</td></tr>
</table>

Figure 3.2: Evidence of whether the GCA-Theorem norm was followed or breached in each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	29001	29002	29003
	Trapezoid	29011	29012	29013
	Isosceles-triangle	29021	29022	29023
	Parallelogram	29031	29032	29033

Figure 3.3: Representation of how each storyboard belongs to one of three conditions and follows one of four storylines

3.1.2. The Items

There are a total of ten items in each questionnaire. They include three items about the first segment of the storyboard, a pair of items about the problem that the class is working on, three items about the second segment of the storyboard, and a pair of items about the teacher's facilitation of the work on the problem throughout the storyboard. The items associated with the first and second segments of the storyboard are the following:

1. What did you see happening in this [first/second] segment of the scenario? (Open response)
2. How appropriate were the teacher's actions in this [first/second] segment of the scenario? (Closed response, specifically, a 6-point Likert item with response options "1–very inappropriate" to "6–very appropriate")
3. Please explain your rating. (Open response)

The pair of items presented after the third segment of each storyboard asks the participant, "How appropriate was the teacher's facilitation of the work on the problem throughout the scenario?" (another 6-point Likert item with response options "1–very inappropriate" to "6–very appropriate"), then asks them to explain their rating. The pair of items about the GCA problem is presented after the participant responds to the three items about the first segment of the storyboard, but before they are presented the second segment of the storyboard. This itemset presents the participant with an image of the GCA problem that the student at the board was solving beside an image of another GCA problem that either follows or breaches the GCA-Figure norm (depending on the storyline), but is otherwise equivalent. The first item asks the participant to compare the appropriateness of the two problems—asking, “Which of the two problems below is more appropriate for a Geometry teacher to present to students?”— using a 6-point scale with response options from “1–Task A is much more appropriate than task B” to “6–Task B is much more appropriate than task A”). The second item asks the participant to explain their rating. This form of this itemset is inspired by the DRN instrument (Herbst Dimmel, & Erickson, 2016; Herbst, Kosko, & Dimmel, 2013), described in the previous chapter. Figure 3.4 provides an example of one of these itemsets.

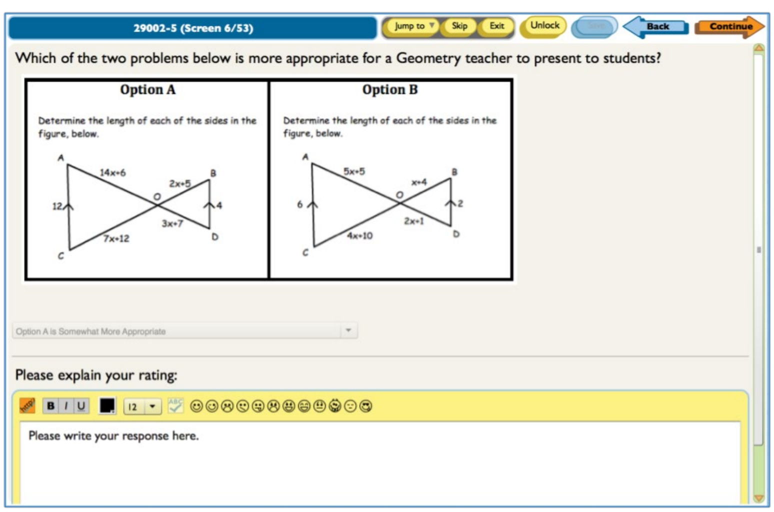


Figure 3.4: Sample DRN-style itemset

The intended design of these itemsets was for the participant to compare the problem in the storyboard to one that followed the GCA-Figure norm in all twelve questionnaires. However, due to an error in these itemsets in the trapezoid questionnaires, the comparison problem breaches the GCA-Figure norm. Consequently, in the compliance-compliance and compliance-breach versions of the trapezoid questionnaire, the itemset asks participants to compare a problem that follows the norm to one that breaches it. In the breach-breach version of that questionnaire, the itemset asks participants to compare a problem that breaches the norm to another problem that breaches it. As such, the responses to this itemset in the three trapezoid questionnaires will be analyzed separately from the responses to this itemset in the other nine questionnaires.

In summary, Figure 3.5 represents how these ten items are sequenced with respect to each other and the three segments of each storyboard. For another perspective, the reader is encouraged to review the images of each of the screens in questionnaire 29002, included in Appendix B. A detailed explanation of how these items allow one to answer the three research

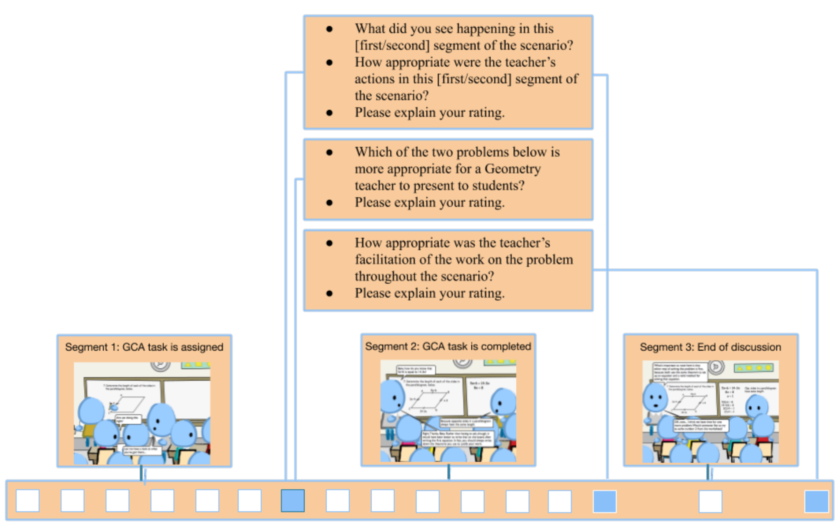


Figure 3.5: Representation of the structure of each questionnaire

questions is provided in the sections of this chapter on the coding of the open-response data and statistical analysis (sections 3.3.1 and 3.3.3, respectively). However, to provide some rationale behind the elements of the design of the instrument just described, it is worth noting here that the "what did you see happening" questions were intended to capture whether participants remarked the decisions to breach or follow each norm (as well as the teacher's request for the students to state or write the geometric property that warrants the first equation in their solution), as well as any attitudes towards such decisions that they may voluntarily express. As mentioned in the previous chapter, the choice to remark decisions to breach a hypothesized norm—in particular, when such remarks are made by experienced geometry teachers, given their familiarity with the instructional situations of the course, and when such remarks are more common than remarks that the hypothesized norm was followed—is evidence that the hypothesized norm does, in fact, exist. This inference is supported by the fact that normative behaviour (being tacitly expected, by definition) tends to go unnoticed, while non-normative behaviour (being unexpected) tends to attract attention (Mehan & Wood, 1975, p.23). Such an inference is also supported by the fact that each storyboard contained other things that a participant might deem noteworthy (described in section 3.1.1) and that the question is worded openly enough to be read as equally inviting of such comments.

In terms of research question 2, evidence from this instrument of experienced geometry teachers' attitudes towards a decision to breach a hypothesized norm and attitudes towards a decision to follow it comes in two forms: evaluations of those actions in participants' open responses to storyboards in which a given norm is breached and ratings of the appropriateness of the teachers' actions in those storyboards. In terms of research question 3, this instrument differs importantly from Dimmel's in that each storyboard represents decisions to follow or breach each of *two* hypothesized norms. It is this feature that allows the researcher to explore whether experienced geometry teachers are more likely to expect the GCA-Theorem norm to be followed, or deem breaches of it (un)acceptable, when the GCA-Figure norm is followed. These relationships are assessed by comparing how experienced geometry teachers respond to storyboards that breach both norms (i.e., breach-breach storyboards) and storyboards that breach the GCA-Theorem norm but follow the GCA-Figure norm (i.e., compliance-breach storyboards). In particular, to assess whether they are more likely to expect the GCA-Theorem norm to be followed when the GCA-Figure norm is followed, one would compare the proportion of experienced geometry teachers that remark a breach of the GCA-Theorem norm across storyboards that breach both norms and ones that only breach the GCA-Theorem norm. To assess whether their attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed, one would compare, across those two types of storyboards, (1) the proportion of experienced geometry teachers that negatively evaluated the breaches of the GCA-Theorem norm, (2) the proportion of experienced geometry teachers that positively evaluated the breaches of the GCA-Theorem norm, and (3) experienced geometry teachers' average rating of the teachers' actions in the second segment of each storyboard.

Last, the following are the names of each item in the dataset, which will be used in some of the discussion of the results and important to anyone conducting further analysis of the data:

- 290xy_1: What did you see happening in this first segment of the scenario?
- 290xy_2: How appropriate were the teacher's actions in this first segment of the scenario?
- 290xy_2R: Please explain your rating.

- 290xy_3: Which of the two problems below is more appropriate for a Geometry teacher to present to students?
- 290xy_3R: Please explain your rating.
- 290xy_4: What did you see happening in this second segment of the scenario?
- 290xy_5: How appropriate were the teacher's actions in this second segment of the scenario?
- 290xy_5R: Please explain your rating.
- 290xy_6: How appropriate was the teacher's facilitation of the work on the problem throughout the scenario?
- 290xy_6R: Please explain your rating.

Note that the root of these variable names, 290xy, is the name of the storyboard with which the item is associated (or, equivalently, the questionnaire in which the item is included), introduced in the previous section.

3.2. Data Collection

Between March 2015 and January 2016, the INR GCA instrument was administered to a national sample of U.S. high school mathematics teachers,[36] along with several other questionnaires, including measures of other constructs (beside norms) that Herbst and Chazan (2012) suggest influence mathematics teachers' instructional decisions, such as the mathematical knowledge needed for teaching geometry (Herbst & Kosko, 2014). The initial design of the sample consisted of two stages: a selection of schools then a selection of one teacher from each school. After obtaining a list of U.S. public secondary schools from the 2012-2013 NCES

[36] Data used in this study was collected with resources from NSF grant DRL- 0918425 to Patricio Herbst. All opinions ensuing from the analysis of this data are those of the author and do not necessarily represent the views of the foundation.

Common Core of Data Public Elementary/Secondary School Universe Survey[37] (our sampling frame for schools), obtaining a list of U.S. secondary school students and teachers from the MCH database (our sampling frame for teachers; www.mchdata.com), and matching students and teachers to schools, we[38] stratified the list of schools by urbanicity[39] and by geographical region[40] and randomly selected 1000 schools from within each urbanicity-by-region stratum with probability proportional to the number of students in that stratum. The second stage consisted of randomly selecting one mathematics teacher from each school and sending them an email invitation to participate in the study (using the email addresses provided by MCH), which explained that each participant would receive $300 in exchange for completing all of the questionnaires, which we estimated would take approximately 24 hours.[41] We chose to email 1000 schools as we hoped to obtain a sample of roughly 600 teachers and because Hill (2007), who used a similar technique to recruit a sample of middle school mathematics teachers, reported that 64% of the people her group emailed agreed to participate in her study. However, this was overly optimistic: only 4% of the teachers that we emailed agreed to participate. We considered using this rate to resample schools, but realized that this would require us to contact 15,000 public secondary schools, which was more than we had in our NCES sampling frame. Resampling therefore consisted of emailing a mathematics teacher that was randomly selected from each of the schools where we did not yet have a participant (and repeating this procedure until over 700 teachers agreed to participate in the study). Nonetheless, the final sample of over 700 teachers that completed the background questionnaire resembled the teachers in the

[37] Actually, similar to how Hill (2007) defined middle schools as "schools that had at least 10 students in each of the sixth and seventh or seventh and eighth grades" (p.10), we defined secondary schools as schools with at least 10 students enrolled in the ninth and tenth or tenth and eleventh grades.

[38] I use the term "we" here because I was not the one who designed the sample, but consulted on its design.

[39] The values of this NCES variable are city, suburb, town, rural.

[40] The values of this MCH variable are Northeast, Southwest, West, Southeast, and Midwest.

[41] Participants earned a $50 gift card at each of 6 checkpoints.

sampling frame (teachers in the MCH database who completed the NCES sample) in terms of the urbanicity and geographical region of the schools that they worked in. They also resemble the teachers in the sampling frame in terms of the number of students eligible for free or reduced-price lunch[42] (a proxy of average student household income) at their schools as well as in terms of their gender and race (Herbst et al., 2017). Out of that sample, 480 teachers completed the four INR GCA questionnaires that they were assigned to complete. 303 of them had three or more years of experience teaching U.S. high school geometry. Given that prior literature suggests that they would be reliable informants on norms of the instructional situations of that course (e.g., Herbst et al., 2013; Herbst, Nachlieli, & Chazan, 2011), their responses from these will be the focus of the analysis.

All participants were randomly assigned to three groups that were each strategically assigned the following four questionnaires:

- Group 1: 29001, 29021, 29012, 29032
- Group 2: 29002, 29022, 29013, 29033
- Group 3: 29011, 29031, 29003, 29023

In other words, group 1 completed two compliance-compliance questionnaires, one in which the storyboard followed the similar-triangles storyline then one in which the storyboard followed the isosceles triangle storyline. Then, they were assigned two compliance-breach questionnaires, one in which the storyboard followed the trapezoid storyline then one in which the storyboard followed the parallelogram storyline. Group 2 was assigned two compliance-breach questionnaires, again one in which the storyboard followed the similar-triangles storyline then one in which the storyboard followed the isosceles storyline. Then, they were assigned two

[42] The values of this NCES variable are over 75% of students, between 50% and 75% of students, between 25% and 50% of students, and less than 25% of students being eligible for free or reduced-price lunches.

breach-breach questionnaires, again one in which the storyboard followed the trapezoid storyline then one in which the storyboard followed the parallelogram storyline. Last, group 3 was assigned two compliance-compliance questionnaires, one in which the storyboard followed the trapezoid storyline then one in which the storyboard followed the parallelogram storyline. Then, they were assigned two breach-breach questionnaires, one in which the storyboard followed the similar-triangles storyline then one in which the storyboard followed the isosceles triangle storyline. While the assignment of participants to groups was random, the size of each group was different: Group 1 included 131 participants (79 of whom had three or more years of experience teaching U.S. high school geometry), while group 2 included 180 participants (107 of whom had three or more years of experience teaching U.S. high school geometry), and group 3 included 182 participants (121 of whom had three or more years of experience teaching U.S. high school geometry). However, as explained in the statistical analysis section of this chapter, the type of regression models that I used to compare scores across questionnaires (mixed-effects models) do not require that groups being compared are equal in size. They also account for the fact that each participant responded to multiple questionnaires.

Last, there was one issue in administration worth noting here: The version of questionnaire 29013 administered to a subset of group 2[43] contained a single image (the image that was included in the DRN-style itemset[44] in this questionnaire), rather than the second segment of the breach-breach version of the similar-triangles storyboard, was included in the first open-response item about the second segment of the storyboard (29013_4). The responses from the participants that received this version of the item were therefore coded as missing, similar to

[43] The erroneous item was administered to 52 of the 173 participants that took questionnaire 29013.

[44] That is, the itemset modeled after the DRN instrument that asked participants to compare two GCA problems, included in each questionnaire.

the responses to open-ended items that contained no words (a system error, as those responses were intended to be forced-response) and that were equivalent to "no comment" (e.g., "nc", "no additional comments"). I did not impute the resulting missing values (in the coded open responses), because the regression models that I used to compare scores across questionnaires are robust against data that is *missing at random* (as this data is), because there were enough participants that took the item after it was corrected to not be concerned about statistical power, and (again) because these models do not require that groups being compared are equal in size.

3.3. Analysis

The analysis of the data consisted of three main stages. First, the responses to the open-ended items were coded. Next, various scores associated with each questionnaire were created from those coded responses (e.g., a score that represents whether any of a participant's open responses to a given storyboard contains evidence that they noticed that the GCA-Figure norm was breached). Last, various statistical models were used to compare average scores and ratings of the teachers' actions in each storyboard (e.g., their rating of the teacher's action in the first segment of a given storyboard) across questionnaires, with particular focus on estimating differences across experimental conditions. In the next three sections, I elaborate on each of these three stages of the analysis.

3.3.1. Coding

The coding of the open-response data also consisted of stages. First, another mathematics education researcher[45] and I coded a stratified random sample of 2,160[46] of the approximately 11,700 responses to the open-ended items. Next, I identified a set of search terms that would allow me to identify all of the responses to which one or more of the codes should likely be applied, based on the reconciled coding of two-thirds (1,440) of the responses that I and the second coder coded together. I then tested them on the reconciled coding of the other one-third (720) of the responses that we coded together. Last, I selected the subset of those search terms that were most efficient in identifying responses to which one or more of the codes should be applied (in the sense that they identified the highest number of such responses and few false positives), applied them to the remainder of the approximately 9,540 responses, then manually coded each of the responses that they identified.[47]

I only asked the second coder to code a subset of the data, because it would have been very costly—both financially and in terms of time—to have had them code all of the responses and meet to reconcile all the differences. I chose to develop and use a set of search terms because it would expedite the process of coding the remainder of the responses, but also because it would extend our understanding of how teachers communicate their recognition of actions that breach or follow the GCA-Figure and GCA-Theorem norms. The potential cost would be a higher

[45] A brief note on the selection of a second coder: While I did not require that they have prior knowledge of the mathematical topic of my research (GCA) or any of the theory (e.g., understanding of norms or linguistics) that undergirded the design of the INR-GCA instrument, I believed (and continue to believe) that it was important that they had enough knowledge of mathematics. In this case, the second coder that I chose had an undergraduate degree in mathematics and a Ph.D. in mathematics education. I argue that this knowledge is important because it is required to interpret some of the responses. For example, I had not originally included the work "invalid sides" as an example of how one might describe a breach of the GCA-Figure norm in my codebook, but both the second coder and I were able to recognize it as such.

[46] These were the set of responses to the 6 open-ended items in each of the 12 questionnaires written by a random 30 of the participants that responded to each questionnaire.

[47] While I am not yet trained in machine learning, some of its basic principles are what guided this approach: developing an algorithm (in this case, a set of search terms that were used to identify responses that I would subsequently read carefully and code manually) based on one dataset, testing it on another, then using the algorithm to analyze a third large dataset.

number of false negatives (i.e., responses to which a code was not applied but should have been) than there would have been if I had manually coded all of the data. That said, there is no reason to believe that the number of such false positives would be unequal across questionnaires (in particular, those belonging to different experimental conditions), and so I argue that it is fair to expect that the results of the statistical tests conducted are similar to what I would have found if I had coded all of the data manually.[48]

My work with the second coder began with me training him on the coding scheme that I created for the pilot study (Boileau & Herbst, 2015, 2016). That coding scheme was based on how the GCA-Figure and GCA-Theorem norms were followed and breached in each storyboard as well as on coding schemes that were developed to analyze open-responses to storyboards in other INR instruments designed to investigate norms of other instructional situations (e.g., Buchbinder, Chazan, & Capozzoli, 2019; Dimmel & Herbst, 2017; Herbst, Aaron, Dimmel, & Erickson, 2013). As mentioned in the previous chapter, these coding schemes included codes that indicate whether a given response contains evidence that the participant noticed that the hypothesized norm(s) under investigation was (or were) breached or followed. These codes were based on an understanding of norms as implicit behavioural rules that people familiar with them tacitly expect others to follow, which render norm-compliant behaviour typically unnoticeable and actions that deviate from them necessary to remark (Herbst & Chazan, 2011; Mehan & Wood, 1975). As breaches of norms are also often sanctioned (Bicchieri, 2006; Garfinkel, 1963) these coding schemes also included codes that indicate whether the decision to breach or follow that norm were deemed acceptable or unacceptable. The original coding scheme therefore

[48] In contrast, I would not suggest that the sample means and proportions that I calculated are representative of their corresponding population parameters. Accordingly, all of the claims made in this dissertation are about relationships between variables (e.g., the choices to remark a given type of action and the experimental condition of the storyboard that the participant is responding to).

included four *recognition codes* that indicate whether the participant remarked that the GCA-Figure norm and/or GCA-Theorem norm was breached or followed,[49] as well as two codes associated with each recognition code that indicate whether the participant deemed any of these actions that they remarked acceptable or unacceptable.

The training began with a meeting in which I introduced the second coder to the work ahead of us: the goals of the study, the norms that I am investigating, and the basic design of the instrument. I then gave them a codebook to review before our next meeting that contained a definition of each code, accompanied by one or more examples of responses to which it should be applied. The plan was to independently code subsets of the data, and meet to reconcile differences in our coding between coding assignments, until we had coded approximately 5% of the data. Then, we would independently code another approximately 5% of the data and calculate our interrater reliability. However, when coding the first subset of the data—720 responses[50]—I realized the need to improve the definitions of some codes, to add new coding rules, and even to develop new codes. I realized how I could improve the definitions of the codes when the second coder would ask questions about the definitions that I originally provided during our meeting. I realized the need to add new coding rules when our discussions of how we coded particular responses revealed that we were each developing our old rules to consistently code aspects of responses that were not addressed by the initial rules, but that seemed quite reasonable when we discussed them. For example, I was coding responses such as the following as containing evidence that the participant noticed that the student had both verbally stated and written the justification for the first equation in their solution:

[49] As mentioned in the previous chapter, remarks of norm-compliant behaviour are coded to help specify the norm—to suggest what actions tend to go unnoticed (because they are expected), in addition to which actions attract attention (because they breach participants' expectation).

[50] The set of responses to the 6 open-ended items in each of the 12 questionnaires written by a random 10 of the participants that responded to each questionnaire.

Verbally stating the theorem was sufficient in my opinion.

As Systemic Functional Linguistics (SFL) encourages the analyst to consider the context of each unit of text that they analyze (Eggins, 2004, p.9)[51] and this was a response to a storyboard in which that justification was both verbally stated and written, I argued that the above response was a statement that writing the justification, or asking for it to be written, was excessive. After some discussion, the second coder agreed with my interpretation of the response, which led me to add more examples of responses to which each appraisal code would apply as well as more explanation of SFL to the codebook.

In terms of adding new codes to the coding scheme, the original coding scheme had included a code that indicated whether a given response contained evidence that the participant that wrote it recognized that the GCA-Theorem norm was followed. This code was to be applied when a response contained evidence that (a) the participant recognized that the student whose work they had seen when viewing the first segment of the storyboard did not include a justification of the equation that the student set up and solved, (b) the participant recognized that the student at the board did not initially justify the equation that they had set up and solved, (c) the participant recognized that the student at the board later verbally stated the geometric property that justified the equation that they had set up and solved, or (d) the participant interpreted the teacher's request for that justification as a request for the student to verbally state it. The code did not distinguish between a student's verbal statement of that justification and the

[51] Eggins (2004) explains that "[o]ur ability to deduce context from text, to predict when and how language use will vary, and the ambiguity of language removed from its context, provide evidence that in asking functional questions about language we must focus not just on language, but on language in context. (p.9)

teacher's request that it be verbally stated because the experiment was designed to test a claim about mode of communication, not division of labour: That a student could be expected to verbally state the property that warrants the equation that they set up to solve a GCA problem, but would not be expected to write it. Similarly, the code did not distinguish between the student not justifying that equation and them verbally justifying it because both actions were hypothesized to be normative (in the specific sense that they were both allowed by the GCA-Theorem norm). However, after coding several responses, the second coder and I noticed that it was uncommon for participants to remark that a student did not justify the first equation in their solution, but fairly common for participants to remark that the student at the board justified that equation or that the teacher asked them to do so. Since normative behaviour tends to go unnoticed (because it is tacitly expected), the fact that it was fairly common to note the verbal justification (and request for it) made me wonder whether the GCA-Theorem norm might be different than I had originally hypothesized: It could be, instead, that the student is only expected to document their algebraic work, making any justification of the equation in their solution—written or verbally stated—a breach of that norm. Again, if this were the case, then a teacher's request for that justification—to be written or verbally stated—would also be non-normative. While the experimental conditions would not allow me to test whether that is the case, curious about this possibility, I decided to break the code indicating whether a given response contained evidence that the participant recognized that the GCA-Theorem norm was followed into two codes: One that would indicate whether the participant recognized that a student did not justify the first equation in their solution and a second that would indicate whether the participant recognized that the student verbally justified that equation and/or that the participant interpreted the teacher's request for that justification as a request for the student to verbally state it. In order to answer the second research question, I also added four related codes that would indicate whether the participant positively and/or negatively appraised any of those actions (the lack of justification, the verbal statement of the justification, or what some participants interpreted as the teacher's request that it be verbally stated).

Another example of a code that was added to the coding scheme was inspired by our realization that it was quite common for participants to remark that the student at the board had justified their first equation and/or that the teacher had asked them to do so, but to not remark the mode of communication used by the student or expected by the teacher (i.e., that the student had verbally stated or written the justification and/or that the teacher had asked them to verbally state or write it).[52] I therefore added a code that would indicate whether a response contains such evidence. I also added codes that would indicate whether a participant positively or negatively appraised either the student's decision to provide that justification and/or the teacher's decision to request that justification in their response to a given open-ended item.

A third example of a code that was added to the coding scheme was inspired by our realization that some participants remarked that the student whose work they see when viewing the first segment of the scenario did not justify their first equation and/or that the student at the board did not initially justify their first equation (until the teacher asked them to). We also noticed that some participants negatively appraised the fact that one or both of those solutions did not include that justification. I therefore also added a code to the coding scheme that would indicate that a participant remarked that the work shown in the first segment of the storyboard did not include that justification or that the work on the board in the second segment did not

---

[52] There are at least two reasons why a participant might offer such a response. The first is that people are not always so specific. Another, which is more theoretically interesting, is that participants were so surprised that a justification was provided or requested at all that the modality of the justification did not stand out to them.

initially contain it, as well as codes that would indicate whether a participant positively or negatively appraised either decision[53] when responding to a given item.

Last, something that became apparent when coding those 720 responses was that participants would not only remark that a norm was breached or followed, as well as whether they approved or disapproved of such decisions, but would also sometimes note conditions under which they would deem breaches acceptable. Consequently, after coding that data using the original coding scheme, we decided to conduct a thematic analysis (Braun & Clarke, 2006) of the same data to develop codes that would allow us to calculate the frequencies with which such conditions were verbally stated. More specifically, first, we each open-coded the data independently. We then met to discuss similarities and differences in our initial codes and negotiate a set of focused codes. One of those focused codes would indicate that the participant suggested that a decision to breach the GCA-Figure norm is acceptable on the condition that the class would discuss the existence of the figure (e.g., what they should take from one of the determined side lengths being 0). The other would indicate that the participant suggested that, if a teacher wants to ask a student at the board to justify the first equation in their solution, they should make this request either before the student starts solving the problem or after they finish solving it, rather than interrupt their work to do so.

Once I had made those revisions to the coding scheme and codebook, the second coder and I each independently applied the revised coding scheme to another 576[54] responses, meeting each week to reconcile differences and negotiate any final small changes needed to the codebook. Last, we each used the revised coding scheme to independently code the remaining

[53] While we only observed a participant negatively appraising either of those decisions in this subset of the data, it was important to also include a code to indicate whether a participant positively appraised either of those decisions in case other responses that I would code later contained such appraisals.

[54] The set of responses to the 6 open-ended items in each of the 12 questionnaires written by a random 8 of the participants that responded to each questionnaire.

1,584[55] responses (making the total number of responses to which the revised coding scheme was applied by both coders 2,160 responses, as mentioned earlier). In the next section, I provide an overview of the final coding scheme along with examples of responses to which each code should be applied.

Overview of the Coding Scheme. The coding scheme consists of three types of codes: *recognition codes*, *appraisal codes*, and *condition codes*. The recognition codes were used to indicate whether a response contains evidence that the participant who wrote it recognized that the GCA-Figure norm and/or GCA-Theorem norm were/was breached or followed in the storyboard to which they were responding (e.g., whether the participant remarked that the GCA-Figure norm was breached or followed, and whether they remarked that the student justified the first equation in their solution). The appraisal codes were used to indicate whether a response contains evidence that the participant deemed any of those decisions acceptable or unacceptable. I refer to these as appraisal codes because their definitions are based on Martin and White's (2005) Appraisal Theory (introduced after the recognition codes). Last, the condition codes were used to indicate whether a response contains evidence that the participant who wrote it thought that the breach of the GCA-Figure norm and/or GCA-Theorem norm represented in the storyboard to which they are responding would be acceptable under certain conditions.

Recognition Codes. There are six recognition codes in total: two related to the GCA-Figure norm and four related to the GCA-Theorem norm. One of those codes (R_GCAF_B) indicates whether a given response contains evidence that the participant that wrote it recognized that the GCA-Figure norm was breached. The following are examples of responses to which this code was applied:

---

[55] The set of responses to the 6 open-ended items in each of the 12 questionnaires written by a random 11 of the participants that responded to each questionnaire.

> The teacher decides to go over a problem from the worksheet on the board. The problem is just a basic parallelogram. The answer the students find is x =1 however that makes **the length of two sides of the parallelogram a negative**. This problem should not be given to students.

> The teacher picked a problem at the request of several students. He gave them time to work on it while a student was preparing to present and checked in on other students during this time. **The side lengths obtained in the end would not make a triangle**.

> The teacher gave the class a choice. They were divided. She compromised. Someone volunteered. A couple minutes were given so everyone could try. **The (impossible) answer** was shared.

> Reluctant student doesn't want to share their **incorrect solution.**

The first three examples may be clear, given the description of the storyboards earlier in this chapter, as they include the terms “the length... [is] negative”, “the side lengths… would not make a triangle” (i.e., the triangle inequality was violated), and “impossible answer”. The fourth response, however, may require some explanation: It is reasonable to interpret a participant’s claim that a student’s solution was erroneous as evidence that they recognized that one or more of the determined side lengths were not possible (i.e., that the GCA-Figure norm was breached) because there were no errors in the solution presented by the students in any of the storyboards. Moreover, the fact that several participants were willing to assume an impossible answer is evidence that the solution is incorrect suggests that the norm is very strong: If it was indeed very rare for the GCA-Figure norm to be breached, then a teacher would be able to reliably infer from an impossible answer to that the solution contains an error.

The second recognition code related to the GCA-Figure norm (R_GCAF_C) indicates whether a given response contains evidence that the participant recognized that the GCA-Figure norm was followed. The following are examples of responses to which this code was applied:

> The solution to the equation **produced a valid side length**.

> Good first example since it deals with **positive numbers.**

> This one is **actually a triangle**.

In terms of the four recognition codes related to the GCA-Theorem norm, the first (R_WJ) indicates whether a given response contains evidence that the participant recognized either the student's breach of the GCA-Theorem norm (i.e., their writing of the property that warranted the equation that they had set up and solved) or the teacher's request for them to breach it. The following are examples of responses to which this code was applied:

> Gamma succeeding and **recording the reason** after prompted to do so and then assuring an incredibly compliant or hesitant student that it's ok to chat…!!!

> The teacher is having the students justify their work by **putting down the theorem** they use to establish the equation. When he asked about the other method the teacher showed the class that sometimes there are more than one solution method.
>
> The teacher allowed the student to do the work, encouraged the student to **list the geometric fact/theorem**, and allowed the students to discuss the solution. However, she does not need to go on to the next problem without talking about whether the solutions made sense or not.
>
> Teacher made student **display why** they were using the defined process for solving.

One thing to note here is the variety of verbs that participants used (and that one might expect high school mathematics teachers to use) to refer to the act of writing other than “write”: “recording”, “putting”, “list”, “display” (to name just a few). Another is that, while it was quite common for participants to refer to the geometric property that warrants the first equation in a student’s solution as a “property”, “theorem”, “justification”, or “reason” it was also quite common to refer to such justification as the reason “why” the problem was solved in the way that they did or the reason “why” the solution is correct. In the next section, I provide the full list of terms used in responses to which one of the codes that indicate that the GCA-Theorem norm was followed or breached or that the teacher asked a student to follow or breach it.

Because one moment in each storyboard when the GCA-Theorem norm is followed is when the student verbally states the geometric property that warrants the first equation in their solution, after being asked to provide such justification, the coding scheme also includes a code (R_SJ) that indicates whether a given open response contains evidence that the participant recognized that the student verbally stated that geometric property and/or interpreted the teacher's request for justification as a request for the student to verbally state that geometric property.[56] The following are examples of responses to which this code should be applied:

> He encouraged the student to **state[57] the theorem** used in the solution.

> Omega started putting her solution up on the board. The teacher asked Omega to **tell why she could set up her equation** and reminded the class that they should always be able to justify why they do things. Omega justified her equation and then finished putting the problem up. Then the teacher asked Alpha to share what she got.

A second way that the GCA-Theorem norm is followed in the storyboards is by a student setting up and solving an equation, but not stating or writing the property that allowed them to set up the equation. As mentioned earlier in this chapter, one place in each storyboard where this happens is in the first segment, when the participant is given an over-the-shoulder view of a complete solution to the problem that does not include a justification for the first equation. Another place is in the second segment of the storyboard, when the student at the board writes an equation and begins to solve it without justifying it, before being interrupted by the teacher who asks them to

[56] Note that I am claiming that this is the participant's interpretation because the teacher in each storyboard asks the student at the board what property they used, but does not specify that they want them to state it verbally. Therefore, when a participant claims that the teacher asked the student to verbally state a geometric property, that is their interpretation of the teacher's question. It is also worth clarifying that this footnote is simply an explanation of what the participant is doing when describing this aspect of the scenario in this way, not an evaluation: Given that the problem is being solved in front of the class, it is reasonable to interpret the teacher's question as a request to verbally state the property. However, especially if it were normative to document justification when solving GCA problems (e.g., as it is when doing proofs in high school geometry; Herbst, Chen, Weiss, & Gonzalez, 2009), it would be fair to interpret the teacher's question as a request for that justification to be written.

[57] It is worth noting that participants would often use expressions like "state the theorem" in response to storyboards in which the geometric property that warrants the first equation in the student's solution was only verbally stated (not also written). Similarly, it was also common for participants to write something like "stated and wrote" when responding to storyboards in which the student both verbally stated and wrote the property. This, in addition to the typical modality of a response to a question during a whole class discussion being verbal, had us apply the R_SJ code to responses that mentioned that the student "stated" the property, and/or that the teacher asked them to "state" it.

justify the equation. The coding scheme therefore includes a code (R_NJ) that indicates whether the participant remarked either of these events. The following is an example of a response to which this code was applied:

> Beta solved the problem but **left out the theoretical foundation that let him create his equation**. If I was the teacher in this class I would be getting quite ticked about this by now as they have been reminded of that necessary step several times.

> The teacher was reviewing the work with the class. **The student left the reasoning off of the problem**, so the teacher made sure the explanation was there. The teacher also noticed another student had worked out the problem differently. The teacher wanted to make sure the students knew there was more than 1 way to work the problem.

> **Gamma is not writing down theorems** so she is stopped by the teacher. Once the problem is solved students are asked to discuss the accuracy of the work Gamma showed.

These quotes reveal that "theoretical foundation" and "reasoning" were amongst the terms that participants used to refer to the justification of the first equation in a student's solution. Also, the terms used to refer to the fact that the justification was not provided: "left out", "left off", and "not writing".

Last, as alluded to earlier, the coding scheme also includes a code (R_JWoM) that indicates whether a given response contains evidence that the participant recognized that the student provided that justification, or that the teacher asked them to, but does not specify whether that justification was written or spoken, or whether the teacher asked the student to verbally state or write it. The following are examples of responses to which this code was applied:

> A student works out a problem. The teacher **requests that she justify a step**. Then they compare the solution to a different student's solution.

> Teacher was asking student to clarify **how they got the left expression equal to the right expression - what geometric properties**

> The teacher **making sure the students know what rule that applies** to trapezoid is being used in order to find the solution.

> The teacher was **reinforcing the mathematical concepts** with the students as the problem was completed. Answers were analyzed for accuracy.

These quotes reveal more terms that participants used (and that one might expect high school mathematics teachers to use) to refer to justification: "how they got the left expression equal to the right" (equivalent to "why she could set up her equation", above), "rule", and "mathematical concept". We also have more examples of terms that participants used to refer to the teacher's request for that justification: "making sure" and "reinforcing".

With the six recognition codes now defined, in the next section, I define the appraisal codes. I begin with a brief introduction to some of the basic principles of systemic functional linguistics and appraisal theory.

Appraisal Theory. Systemic Functional Linguistics (SFL) is "a theory of language as social process and an analytical methodology which permits the detailed and systematic description of language patterns" (Eggins, 2004, p.21). As a theory of language as social process, SFL acknowledges that language evolved and is used to construe meaning and that the meaning that it construes is dependent on the situational and cultural contexts in which it evolved and is used. The latter is particularly relevant to the analysis of data collected with the INR GCA instrument because participants typically refer to aspects of the storyboards that they are being asked to consider, rather than the abstract concepts that they represent (e.g., compliance with a given norm). For example, in the pilot study, one participant responded to a storyboard in which the teacher breached the GCA-Theorem norm by saying, "the kid wrote what he/she just said out loud". While this statement is rather unclear in isolation—What kid? What did they say out loud?—it is quite clear when one considers the storyboard that they are responding to: The only student who writes something that they also verbally state is the student at the board and the only thing that they both verbally state and write is the geometric property that allows them to set up their equation.

Another central idea in SFL is that a text should be understood as a series of choices—some conscious and some unconscious—between what one says or writes and what one could have said or written, as it is these alternatives that give units of text (e.g., words and clauses) their meaning. Tying this idea to the previous one, Schlepegrell (2012) explains that:

> SFL facilitates exploration of meaning in context through a comprehensive text-based grammar that enables analysts to recognize the choices speakers and writers make from

> linguistic systems and to explore how those choices are functional for construing meanings of different kinds. (p.21)

Appraisal is one of those linguistic systems. Appraisal theory—the description of that system—explains “how evaluation is established, amplified, targeted and sourced” (Martin & White, 2005, p.9). Relevant to the present study, the theory accounts for how the choice to appraise something lies in contrast to the decision to simply describe it, and that both lie in contrast to the choice to not comment on it at all. For example, as explained earlier, it is the fact that there are many aspects of each scenario that a participant could comment on that makes it meaningful for one to remark when either hypothesized norm is breached. As norms are behavioural rules that people expect each other to follow, a participant’s choice to remark the breach of a hypothesized norm is specifically interpreted as indication that they expected the norm to be followed. Similarly, given that a participant could simply remark that breach, their choice to appraise it is what makes it possible to infer from their response that a participant holds an attitude towards that behaviour. One step further, it explains that appraisals come in three main forms: expressions of affect, judgement, and appreciation. Martin and White (2005) define, and differentiate between, these as follows: While “[a]**ffect** deals with resources for construing emotional reactions”, “[a]**ppreciation** looks at resources for construing the value of things” and “[j]**udgement** is concerned with resources for assessing behaviour” (p.35). The appraisal codes in my coding scheme do not differentiate among these categories, but they are worth stating here as they help the reader imagine the range of things that could be captured by the positive and negative appraisal codes (e.g., an expression of one’s emotion towards a decision to breach the GCA-Figure norm, an evaluation of the assigned problem, or an evaluation of the teacher’s decision to ask the student at the board to justify the first equation in their solution). With this explained, I can now move on to the definitions of the appraisal codes.

Appraisal Codes. As mentioned earlier, there are a total of twelve appraisal codes: a positive appraisal code and a negative appraisal code associated with each of the six recognition codes. They are associated in the sense that each indicates whether a participant positively or negatively appraised a decision to follow or breach either hypothesized norm (or to ask the student to follow or breach the GCA-Theorem norm). The following are the names of those codes:

- PA_GCAF_C: Positive appraisal of GCA-Figure norm compliance
- NA_GCAF_C: Negative appraisal of GCA-Figure norm compliance
- PA_GCAF_B: Negative appraisal of GCA-Figure norm breach
- NA_GCAF_B: Negative appraisal of GCA-Figure norm breach
- PA_NJ: Positive appraisal of a student's decision to not justify the first equation in their solution
- NA_NJ: Negative appraisal of a student's decision to not justify the first equation in their solution
- PA_JWoM: Positive appraisal of a student's decision to justify the first equation in their solution and/or of the teacher's decision to request that justification, without mention of the modality of the provided and/or requested justification
- NA_JWoM: Positive appraisal of a student's decision to justify the first equation in their solution and/or of the teacher's decision to request that justification, without mention of the modality of the provided and/or requested justification

- PA_SJ: Positive appraisal of a student’s decision to verbally state the justification of the first equation in their solution and/or of the teacher’s decision to verbally state that justification
- NA_SJ: Positive appraisal of a student’s decision to verbally state the justification of the first equation in their solution and/or of the teacher’s decision to verbally state that justification
- PA_WJ: Positive appraisal of a student’s decision to write the justification of the first equation in their solution and/or of the teacher’s decision to write that justification
- NA_WJ: Negative appraisal of a student’s decision to write the justification of the first equation in their solution and/or of the teacher’s decision to write that justification.

The following are examples of responses that would be coded as a negative appraisal of a decision to breach the GCA-Figure norm (NA_GCAF_B), a positive appraisal of a decision to breach the GCA-Figure norm (PA_GCAF_B), and a positive appraisal of a decision to follow the GCA-Figure norm (PA_GCAF_C), respectively:

> Wow I chose some **bad** numbers for the original sides

> Option B presents an issue because the bottom turns out to be -4 which is arithmetically sound (the top is 22 and the median is 9) but geometrically impossible. Still, though, it's a **worthwhile** example to discuss as a class for precisely that reason!

> It was a **good** problem. The solution to the equation produced a valid side length.

The following is an example of a response that contains both a negative appraisal and a positive appraisal of the GCA-Figure norm, exemplifying that a given response could receive both appraisal codes associated with a given recognition code:

> The teacher is having students attempt a trapezoid problem finding the bases and the median. Though the student does the problem correctly, the side length comes out as 0, in which case, we do not have a trapezoid. The expressions **should have been** checked prior to attempting the problems, **however, a discussion about impossible figures could ensue**.

This example is also important because it demonstrates how conjunctions (e.g., “however”) can be important to interpreting clauses: While the statement that “a discussion about impossible figures could ensue” sounds optimistic about that discussion and might be coded as a positive appraisal even if that was all the participant wrote, the fact that it is contrasted with the earlier negative appraisal (“should have”) makes the meaning of the final clause clearer.

The following response is an example of a positive appraisal of the decision made by the student at the board to justify the first equation in their solution that does not specify the expected modality of that justification (PA_JWoM):

> It is **crucial** that students are justifying their work.

The following responses are examples of positive appraisals of the teacher's request that the student at the board justify the first equation in their solution that does not specify the expected modality of that justification (also captured by the PA_JWoM code[58]):

> I **like** the questioning, "What theorem did do you use?"

> Leading questions and request for justification is **part of the mathematical process**.

> The teacher had the student justify their reasoning, which allowed **the teacher to check for understanding** and for the **other students** who didn't get it **to possibly better understand**.

The first of these three responses and the example of an appraisal of the student's decision to justify the first equation in their solution are examples of appraisals that one may be familiar with from casual conversations: "like" and "crucial". In contrast, the next two examples require an understanding of the work of teaching. First, the claim that justification is part of the mathematical process is a way of saying that the teacher in the storyboard attended to one of their professional obligations: to "steward a valid representation of the discipline of mathematics" including its "mathematical practices" (Herbst & Chazan, 2011, p.450). Similarly, it mentions that the students may learn from a given teacher decision that can be interpreted as a positive appraisal because the teacher's job is also responsible for helping all students understand the course content.

[58] As a reminder, the codes that indicate whether this justification is remarked or appraised do not differentiate between comments on the student's decision to provide that justification and the teacher's request for the student to do so, as the teacher's decision to request for justification would be normative if and only if the decision to provide that justification were normative.

The following responses are examples of a positive appraisal of the teacher's request for a student to verbally justify the first equation in their solution (PA_SJ) and a negative appraisal of the teacher's request for a student to write that justification (NA_SJ), respectively.

> I think that it was **good** that the teacher had the student working at the board explain why they were solving the problem the way that they were solving it.

> **I don't think I would have** forced the student to write down the reason he solved the problem if he could just explain it in words.

One thing that is noteworthy about the first example is the evidence that the teacher is referring to a request for the justification to be *stated*, as this is another place where the context is relevant. While a teacher may include a request for a student to explain their thinking on a written test and expect that the student would write it, given that the participant is responding to a representation of a whole class discussion, it is likely that they wrote "explain" because they interpreted the teacher's request for justification as a request for that justification to be stated. The second example may be clearer, but it may also be worth noting that one way that people suggest that they disagree with someone's decision is by explaining that they would not have done that.

Finally, the following responses are examples of negative appraisals of the student's decision not to justify the first equation in their solution (NA_NJ):

> **Never heard [of] the need to** state the theorem when creating an equation???

> Leaps of reason are **not clear to everyone** and **should be** explained.

The appraisal in the first of these examples—"never heard of"—is a common expression in English (at least in American English), but that I include here for those unfamiliar with it. It implies that the speaker expected that something should have been done, but was not done. In this case, the participant seems to have expected the student at the board to verbally state the justification of the first equation in their solution. The second appraisal in the second example—"should be"—is similar. The first appraisal in that example is related to the notion that the teacher is responsible for helping all students understand the course content (mentioned above).

Condition Codes. Finally, as mentioned earlier, there are two condition codes. One of them (D_GCAF_B) indicates that the participant suggested that, if the GCA-Figure norm is breached, the existence of the figure should be discussed. The other (WRJ) indicates that the participant suggested that, if the teacher wants to ask the student at the board to justify the first equation in their solution, they should make this request either before the students start solving the problem or after they finish solving it, rather than interrupt their work to do so.

The following responses are examples of responses in which the participant suggests that breaches of the GCA-Figure norm is breached should be discussed:

> I don't feel that board work should be an intimidating situation. I think the teacher should give an indication that something might be wrong first. **Perhaps the teacher would like to address** the negative side lengths**?**

> **No mention of** getting impossible side length.

As noted in the section on the recognition codes, it was common for participants to remark the breach of the GCA-Figure norm by remarking its consequences—negative, or otherwise impossible, side lengths—as these two participants did. In the first example, the participant suggests that the teacher should have responded to, or had the class discuss, the impossible solution by calling into question the fact that they did not.[59] The logic behind claiming that the second statement is evidence that the participant expected the breach of the GCA-Figure norm to be discussed is the same as the logic behind the recognition codes: If the participant did not expect this to be discussed, they would not have said that it was not mentioned.

The following responses are examples of responses in which the participant suggested that the teacher should have asked students to justify the equations that they set up towards solving the problem either before the students start solving the problem or after they finish solving it, rather than interrupt their work to do so:

> **Teacher should have the student write the entire problem** before interruption **or tell the student before she starts** to make sure to include the theorem.

> Once again, I think **it would be better if the teacher waited until Gamma had done her algebraic calculations before asking her to write the property**. She could have also asked the class "Did Gamma forget anything?"

Something that is noteworthy about both examples is that they only suggest that the teacher should have waited until the student was finished solving the problem before asking for written

[59] As a reminder, in each scenario, after the teacher asks the students to discuss the proposed solution in their groups, they move on to another problem.

justification. It is therefore possible that these participants deemed it acceptable to interrupt to ask for the property to be verbally stated and only needed to wait to insist that the justification be written. However, this code does not distinguish between modality. Instead, it is applied to suggestions that the request for that justification to be written should have come at a different time, suggestions that the request for that justification to be verbally stated should have come at a different time, and suggestions that, that the justification should come at a different time that do not specify the modality of that justification. Furthermore, it is applied to all suggestions that a request for justification of the first equation in one's solution should come before the presentation of a solution, suggestions that it should come afterwards, and suggestions that either would be fine. With these codes now defined and examples now given, I move to describing the process of using a set of search terms to locate them.

Coding Using Search Terms. As mentioned earlier, the next stage in the analysis was to determine a set of search terms that could be used to identify responses to which one or more of the codes should likely be applied, which I would then carefully read and manually code. This process would be used to expedite the coding of the remaining approximately 9,540 responses to which the finalized coding scheme had not yet been applied.[60] This stage of the coding began with carefully reading two-thirds (1,400) of the responses that the second coder and I had coded together and to which we agreed at least one of the codes in the coding scheme had been applied. Since all of the codes in the coding scheme are specific to the norms under investigation and I would code each of the located responses, it became clear that the only terms that I would need were descriptors of the decisions to either follow or breach either the GCA-Figure or GCA-

[60] This included recoding the 720 responses to which we had applied the original coding scheme and on which we conducted the thematic analysis that resulted in the condition codes.

Theorem norm; by their definitions, each recognition, appraisal, and condition code would only be applied to responses that contain such descriptors.

That exhaustive list, however, was quite long, as some terms were only used once or twice in all of the 1,440 responses. The following are the full lists of terms related to each norm (excluding typos and plural versions of words, which would capture by using roots of these words):

- Terms associated with the GCA-Figure norm:
    - Positive, negative, zero, inequality, 0, -[61]
    - Exist, im/possible, in/valid, un/usual, un/realistic, routine, extraneous, sense, sensical, in/correct, error, flaw, mistake, un/reasonable, works, doesn't work, does not work
    - Side, segment, length, measure
    - Triangle, parallelogram, trapezoid
    - Actual, actually[62]
    - Poor problem, bad problem
- Terms associated with the GCA-Theorem norm:
    - Justify, justification, theorem, theory, theoretical foundation, property, reason, reasoning, principle, rule, formula, given, assumption, rationale, concept, statement, basis, support, known, fact
    - Why, lead to
    - Math, process, precision, vocabulary, information

[61] While only a couple of the negative numbers were included in the 1,440 responses, I imagined that others were likely to come up in the rest of the data, so decided to list the symbol "-", rather than only the negative numbers that I had observed up to that point.

[62] In case the ways in which these two terms might be relevant is less clear, recall the example of a remark that the GCA-Figure norm was followed, provided earlier: "This one is actually a triangle."

    - Interrupt, intervene, interject, continue, stop, pause, correction, add; when, before, while, as
    - Step, equal, equation, congruent, set
    - geometric, algebraic

These terms were then tested on the remaining 720 of the 2,160 responses. In doing so, it not only became clear that the use of this list of terms was not only inefficient because of its length, but also because some terms were only used once or twice in the entire set of responses (including the 1,440 from which they were derived) and many of them produced mainly false positives. For example, the terms "valid" and "doesn't work" were quite uncommon, while terms like "reasonable" resulted in mostly false positives, as they were used to in appraisals of other decisions (e.g., the teacher's decision to not allow a particular student to solve the problem at the board). Similarly, the terms "principle", "rule", and "rationale" were quite uncommon, while terms like "reason" and "equation" resulted in mostly false positives, as the former was often used to evaluate other decisions made by the teacher (e.g., their decision to ask the students' to evaluate the work presented at the board) and the latter was often used to refer to aspects of the student's algebraic work. For that reason, from each of these two lists, I selected a subset of the terms that had proven to be efficient in the sense that they identified several responses to which one or more codes associated with the given norm were applied and very few responses to which none of those codes were applied. The final list of terms includes the following:

- Terms associated with the GCA-Figure norm: Positive, negative, zero, inequality
- Terms associated with the GCA-Theorem norm: Justification, theorem, property

I then used these terms to locate which out of the remaining approximately 9,540 open responses I should code. A value of 0 for each code was applied to each response not located by these terms, other than missing responses (as mentioned earlier, either blank cells or ones that contained terms like “no comment”, “nc”, and “no additional comments”).

As mentioned earlier, the cost of using these short lists of terms to code the data, rather than the full list of terms or carefully coding all 11,700 responses, is a higher number of responses to which one or more codes would have been applied. However, there is no reason to believe that the proportion of such false positives is greater in the coding of the responses to any of the twelve questionnaires. Therefore, it is reasonable to expect that the differences, between any two questionnaires, in the proportions of responses to which each of the scores based on these codes were applied would be similar to what they would have been if I had used the full list of search terms or coded each response.[63] In the next section, I describe those scores.

3.3.2. Participant Scores Based on the Coding of the Open-response Data

Some of the answers to my research questions that I offer in the next chapter are based on comparing, across questionnaires, the proportion of experienced geometry teachers who offered responses to a specific item in each questionnaire to which a particular code was applied. Accordingly, I created a total of two scores of this type. The first represents whether (1) the code representing that a participant recognized that the GCA-Figure norm was breached (R_GCAF_B) was applied to a participant’s response to the first item in a questionnaire (290xy_1) that contains a storyboard in which that norm is breached, or (2) the code representing that a participant recognized that the GCA-Figure norm was followed (R_GCAF_C) was applied

[63] In contrast, I would not suggest that the sample means and proportions that I calculated are representative of their corresponding population parameters. Accordingly, all of the claims made in this dissertation are about relationships between variables (e.g., the choices to remark a given type of action and the experimental condition of the storyboard that the participant is responding to).

to a participant's response to the same item in a questionnaire (290xy_1) that contains a storyboard in which that norm is followed. This code is named R_GCAF_1 to reflect that it represents that the GCA-Figure norm was remarked when responding to the first item in each questionnaire. The second score of this type represents whether (1) the code representing that a participant recognized that the GCA-Theorem norm was breached (R_WJ) was applied to a participant's response to the first item in a questionnaire associated with the second segment of the contained storyboard (290xy_4) and that the GCA-Theorem norm is breached in that storyboard, or (2) one of the codes representing that a participant recognized that the GCA-Theorem norm was followed (R_SJ or R_NJ) was applied to a participant's response to the same item in a questionnaire that contains a storyboard in which the GCA-Theorem norm is followed, but not breached afterwards. This code is named R_GCAT_4 to reflect that it represents that the GCA-Theorem norm was remarked when responding to the first item in each questionnaire associated with the second segment of the contained storyboard. The value of both of these scores is that they capture a participant's initial reaction to the norm in question being breached (or followed) and, therefore, is arguably the best measure of whether they expected the norm would be followed.

In contrast, when comparing participants' appraisals of decisions to breach or follow each norm, I considered all responses to each storyboard, as it is not relevant whether they immediately appraise any of those decisions or whether we needed to ask a participant to evaluate the actions taken in a storyboard in order to elicit them.[64] Moreover, it is reasonable to expect that some participants may not have realized that the GCA-Figure norm is breached or

[64] To be clear, whether it is common for one to express their attitude when simply asked to describe what they noticed about a storyboard or whether one needs to be asked to evaluate the actions taken in the storyboard to elicit that attitude is an important methodological question. My point here is that there is no reason to believe that participants would be less truthful about their attitudes about being asked several questions or specifically after being asked to evaluate actions taken in a storyboard.

followed until the second time they reviewed the first segment of a storyboard, or even until they consider the second segment of the storyboard in which the solution is presented at the board. Similarly, it is possible that a participant might not notice that the student at the board had justified the first equation in their solution or that the teacher had asked them to until they are asked to evaluate the teacher's actions in the second segment of the storyboard (in items 290xy_5 and 290xy_5R) or their actions throughout the scenario (in items 290xy_6 and 290xy_6R). Or perhaps a participant did not interpret the question of what they saw happening in a given situation as a place to critique, so would not have appraised the teacher's (or student's) decision to follow or breach a norm until they were asked to evaluate the actions of the teacher. For that reason, I created a second type of score that represents whether *at least one* of a participant's responses to a given storyboard received one or more of the appraisal codes related to one of the norms in question. More specifically, I created scores that reflect whether any of the remarks captured by one of the recognition scores just listed included a positive or negative appraisal of the decision in question. These are named PA_GCAF_any, PA_GCAT_any, NA_GCAF_any, and NA_GCAT_any.

With these scores now described, I can move on to describing the statistical models that I used to compare them across storyboards in order to answer my three research questions:

1. Do the GCA-Figure and GCA-Theorem norms exist? That is,
    a. Do experienced geometry teachers expect GCA problems to have geometrically-meaningful solutions?

    b. Do experienced geometry teachers expect students to either only share their algebraic work, or to sometimes also verbally state the geometric properties that warrant the equations that they set up, but not to write those properties?
2. Do experienced geometry teachers deem decisions to breach either norm acceptable or at least as acceptable as decisions to follow them? That is,
    a. Do experienced geometry teachers deem other U.S. high school geometry teachers' decisions to assign problems that breach the GCA-Figure norm acceptable or at least as acceptable as decisions to assign problems that follow it?
    b. Do experienced geometry teachers deem students' decisions to breach the GCA-Theorem norm and/or U.S. high school geometry teachers' decisions to ask students to do so acceptable or at least as acceptable as decisions to follow it?
3. Are experienced geometry teachers more likely to expect that the GCA-Theorem norm will be followed, or to disapprove of breaches of the GCA-Theorem norm, when the GCA-Figure norm is followed (than when it is breached)?
    a. Are experienced geometry teachers' expectations about whether and how students will share the geometric properties that warrant the equations that they set up, or their expectations about whether and how other U.S. high school geometry teachers will ask their students to provide such justification, dependent on whether the problem has a geometrically-meaningful solution?
    b. Are experienced geometry teachers' attitudes towards students' decisions to write down the geometric properties that warrant the equations that they set up, or other U.S. high school geometry teachers' decisions to ask them to do so, dependent on whether the problem has a geometrically-meaningful solution?

3.3.3. Statistical Analyses

As mentioned above, when conducting a virtual breaching experiment, one concludes that a hypothesized norm exists when a sufficiently large proportion of participants remark that the hypothesized norm was breached in one or more of the representations of the instructional situation to which the norm applies (e.g., storyboards) that they are presented. However, the proportion of population that expect a given norm to be followed is typically unknown, as this can vary considerably, for example, depending on the size of the population and how long the norm has existed (Bicchieri, 2006, p.12; see also Gibbs, 1965). For that reason, in a virtual breaching experiment with controls (such as this one), the proportion of participants that remark the breach(es) of a norm is compared to the proportion of participants that remark that it is followed in storyboards that differ only in terms of whether that norm is breached or followed. Similarly, to determine whether decisions to breach the norm in question are deemed as acceptable as decisions to follow it, the proportion of participants that positively or negatively appraised decisions to breach the norm is compared to the proportion of participants that positively or negatively appraised decisions to follow the norm. For the same reason, the same is done with participants' ratings of the teacher's actions in segments in which the norm is followed or breached in each storyboard. Therefore, the principal role of the statistical analyses that I conducted was to compare the scores described in the previous section as well as the responses to closed-ended items (ratings of the teacher's actions) across questionnaires, particularly across questionnaires that contain storyboards that belong to different experimental conditions, but follow the same storyline. For example, to assess whether the GCA-Figure norm exists, I compared the proportion of experienced geometry teachers[65] who remarked that the GCA-Figure norm was breached when responding to the first item associated with the first segment of each breach-breach storyboard (in which the GCA-Figure norm is either followed or breached) to the proportion of such participants who remarked that the GCA-Figure norm was followed when responding to that item in the compliance-compliance and compliance-breach version of that questionnaire.[66] The reason for comparing that segment of storyboard in those pairs of conditions is that the first segment of each breach-breach storyboard differs only from that segment of the corresponding compliance-compliance and compliance-breach storyboard in terms of whether the GCA-Figure norm is breached. Similarly, to assess whether the GCA-Theorem norm exists, I compared the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was breached when responding to the first item associated with the second segment of each of the compliance-breach storyboards to the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was followed when responding to that segment of the corresponding compliance-compliance storyboard,[67] because that segment of storyboards that follow the same storyline differ only in terms of whether the GCA-Theorem norm is breached.

For the same reason, in order to assess whether experienced geometry teachers deem breaches of the GCA-Figure norm acceptable, I compared average appraisal scores associated with the GCA-Figure norm,[68] as well as average ratings of the teacher's actions in the first segment of each storyboard, across the breach-breach and compliance-compliance versions of each questionnaire, as well as between the breach-breach and compliance-breach versions of each questionnaire. Similarly, in order to assess whether experienced geometry teachers deem

---

[65] As mentioned earlier, I focused on this subset of the participants as prior literature suggests that they would be the most reliable informants on the norms of instructional situations in U.S. high school geometry.

[66] Using the R_GCAF_1 score

[67] Using the R_GCAT_4 score

[68] PA_GCAF_any and NA_GCAF_any

breaches of the GCA-Theorem norm acceptable, I compared average appraisal scores associated with the GCA-Theorem norm,[69] as well as average ratings of the teacher's actions in the second segment of each storyboard, across compliance-breach and compliance-compliance questionnaires.

Last, in order to answer my third research question, I compared average recognition scores associated with the GCA-Theorem norm, average appraisal scores associated with the GCA-Theorem norm, and average ratings of the teacher's actions in the second segment of each storyboard across breach-breach and compliance-breach questionnaires. Comparing these average scores and ratings across these conditions allowed me to answer this question because the GCA-Theorem norm is breached in the storyboards contained in each, while the GCA-Figure norm is only breached in the storyboards contained in the breach-breach questionnaires.

In terms of how I compared these average participant scores (defined in section 3.3.2) and ratings across questionnaires, I used mixed-effects regression models (using the *melogit* and *mixed* commands in STATA). Each model regressed a given participant score or rating on fixed effects indicating the experimental condition and storyline of the storyboard to which the score or rating is associated, as well as a random intercept indicating the participant that contributed it. This random effect was included to account for the fact that each participant responded to multiple storyboards.[70] Because the research questions are about experienced geometry teachers' expectations and attitudes, each model also included a third fixed effect indicating whether the participant had sufficient experience teaching U.S. high school geometry to expect that they would be aware of the norms of the instructional situation of GCA. Based on prior research (e.g.,

---

[69] PA_GCAT_any and NA_GCAT_any

[70] I am including this variable as a random effect, rather than as a fixed effect, because the participants are a random sample of U.S. high school mathematics teachers and my research questions do not require estimating differences in how each individual responded to each storyboard. It is important to include the variable in the model, however, as the model would otherwise be an OLS regression, which assumed that all observations (in this case, scores or ratings) are independent.

Herbst, Aaron, Dimmel, & Erickson, 2013; Herbst, Nachlieli, Chazan, 2011), I define an *experienced geometry teacher* as a teacher who had taught U.S. high school geometry for three or more years. Each model also includes interactions between all of the fixed effects (condition, storyline, and experience) so that I could estimate how experienced geometry teachers would respond to each questionnaire (e.g., whether they would remark that the GCA-Figure norm was followed or breached or how they would rate the appropriateness of the teacher's actions during the first segment of the included storyboard).

Each of these models can be described algebraically as follows:

$$\begin{aligned}
Y_{ij} = {} & \beta_{0j} + \beta_{1j}C\text{-}B_j + \beta_{2j}B\text{-}B_j + \beta_{3j}Trap_j + \beta_{4j}IsoTri_j + \beta_{5j}Par_j + \beta_{6j}Experience_j \\
& + \beta_{7j}C\text{-}B_j{*}Trap_j + \beta_{8j}B\text{-}B_j{*}Trap_j + \beta_{9j}C\text{-}B_j{*}IsoTri_j + \beta_{10j}B\text{-}B_j{*}IsoTri_j + \beta_{11j}C\text{-}B_j{*}Par_j + \beta_{12j}B\text{-}B_j{*}Par_j + \\
& \beta_{13j}C\text{-}B_j{*}Experience_j + \beta_{14j}B\text{-}B_j{*}Experience_j + \beta_{15j}Trap_j{*}Experience_j + \beta_{16j}IsoTri_j{*}Experience_j + \beta_{17j}Par_j{*}Experience_j \\
& + \beta_{18j}C\text{-}B_j{*}Trap_j{*}Experience_j + \beta_{19j}B\text{-}B_j{*}Trap_j{*}Experience_j + \beta_{20j}C\text{-}B_j{*}IsoTri_j{*}Experience_j + \\
& \beta_{21j}B\text{-}B_j{*}IsoTri_j{*}Experience_j + \beta_{22j}C\text{-}B_j{*}Par_j{*}Experience_j + \beta_{23j}B\text{-}B_j{*}Par_j{*}Experience_j \\
& + u_j + \varepsilon_{ij}
\end{aligned}$$

Equation 3.1: Algebraic representation of the type of regression model used to compare scores across questionnaires

In this equation, Y_{ij} is the particular score or rating (i.e., observation) contributed by a participant. $C\text{-}B_j$ *and* $B\text{-}B_j$ are dummy variables indicating whether the score or rating is associated with a compliance-breach questionnaire and whether the score is associated with a compliance-breach questionnaire (versus a compliance-compliance questionnaire),[71] respectively. $Trap_j$, $IsoTri_j$, and Par_j are dummy variables representing whether the score or rating is from the trapezoid, isosceles-triangle, or parallelogram questionnaires (versus a similar-triangles questionnaire),[72]

[71] When the response is associated with a compliance-compliance storyboard, the value of both of these dummy variables is 0.

[72] When the response is associated with a similar-triangles storyboard, the value of each of these dummy variables is 0.

respectively. *Experience*$_j$ is the aforementioned dichotomous indicator of whether the participant to whom the score or rating is associated (i.e., who wrote the response(s) used to calculate the score is based or who made the rating) is an experienced geometry teacher. Finally, u_j is the aforementioned dichotomous indicator of the participant that offered a given score or rating.

In addition to using these models to compare average participant scores and ratings across questionnaires, I also used tests of proportions to compare the proportion of experienced geometry teachers that responded in one way versus another to a storyboard (using the *prtest* command in STATA). For example, I used this type of test to compare the proportion of experienced geometry teachers that positively appraised the breach of the GCA-Figure norm in each of the breach-breach storyboards to the proportion of experienced geometry teachers that negatively appraised it, in order to assess whether such teachers tend to deem decisions to breach the GCA-Figure norm acceptable or unacceptable. In the next chapter, I present the results of these analyses and discuss the answers they provide to the three research questions.

Chapter 4 - Results

In this chapter, I present the results of the statistical analyses that I conducted in order to answer my three research questions. As a reminder, those questions are the following:

1. Do the GCA-Figure and GCA-Theorem norms exist? That is,
    a. Do experienced geometry teachers expect GCA problems to have geometrically-meaningful solutions?
    b. Do experienced geometry teachers expect students to either only share their algebraic work, or to sometimes also verbally state the geometric properties that warrant the equations that they set up, but not to write those properties?
2. Do experienced geometry teachers deem decisions to breach either norm acceptable or at least as acceptable as decisions to follow them? That is,
    a. Do experienced geometry teachers deem other U.S. high school geometry teachers' decisions to assign problems that breach the GCA-Figure norm acceptable or at least as acceptable as decisions to assign problems that follow it?
    b. Do experienced geometry teachers deem students' decisions to breach the GCA-Theorem norm and/or U.S. high school geometry teachers' decisions to ask students to do so acceptable or at least as acceptable as decisions to follow it?
3. Are experienced geometry teachers more likely to expect that the GCA-Theorem norm will be followed, or to disapprove of breaches of the GCA-Theorem norm, when the GCA-Figure norm is followed (than when it is breached)?

a. Are experienced geometry teachers' expectations about whether and how students will share the geometric properties that warrant the equations that they set up, or their expectations about whether and how other U.S. high school geometry teachers will ask their students to provide such justification, dependent on whether the problem has a geometrically-meaningful solution?
b. Are experienced geometry teachers' attitudes towards students' decisions to write down the geometric properties that warrant the equations that they set up, or other U.S. high school geometry teachers' decisions to ask them to do so, dependent on whether the problem has a geometrically-meaningful solution?

As mentioned in the methods chapter, the statistical model that I used to answer research question 3a is the same as the model I used to answer research question 1b, as both questions can be answered by comparing the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm is followed or breached in a given storyboard across experimental conditions (compliance-breach to breach-breach and compliance-compliance to compliance-breach, respectively) and each regression model allows for comparisons across each pair of conditions. To answer research question 3a, I compared experienced geometry teachers' reactions to the breach-breach and compliance-breach versions of each storyboard; to answer research question 1b, I compared their reactions to the compliance-compliance and compliance-breach versions of each storyboard. Similarly, the statistical model that I used to answer research question 3b is the same as the model that I used to answer research question 2b, since both can be answered by comparing the proportion of experienced geometry teachers that (positively or negatively) appraised the decision to breach or follow the GCA-Theorem norm when responding to a given storyboard (represented by the PA_GCAT_any score and the NA_GCAT_any score, respectively) across experimental conditions. Again, to answer research question 3b, I compare their reactions to the breach-breach and compliance-breach versions of each storyboard; to answer research question 2b, I compare their reactions to the compliance-compliance and compliance-breach versions of each storyboard. For this reason, rather than organize the presentation of the results by research question, I organize them by norm.

More specifically, section 4.1 presents the results related to the GCA-Figure norm. Section 4.1.1 presents the results of the analyses conducted to confirm the existence of that norm (i.e., to answer research question 1a). Section 4.1.2 presents the results of the analyses conducted to explore whether experience geometry teachers deem breaches of the GCA-Figure norm acceptable or at least as acceptable as decisions to follow it (i.e., to answer research question 2a). Section 4.2 presents the results related to the GCA-Theorem norm. Section 4.2.1 presents the results of the analyses conducted to confirm the existence of that norm (i.e., to answer research question 1b), and to explore whether experienced geometry teachers are more likely to expect the GCA-Theorem norm to be followed when the GCA-Figure norm is followed than when it's breached (i.e., to answer research question 3a). Section 4.2.2 presents the results of the analyses conducted to explore whether experienced geometry teachers deem breaches of the GCA-Theorem norm acceptable, or at least as acceptable as decisions to follow it (i.e., to answer research question 2b). It also presents the results of analyses conducted to explore whether their attitudes towards breaches of that norm depend on whether the GCA-Figure norm is breached (i.e., research question 3b). Section 4.2.3, I call into question whether the GCA-Theorem norm is what I originally hypothesized that it was.

A final thing worth noting is how the results of each regression analysis are presented. While it is common to present the values of each of the coefficients in the regression equation (i.e., parameter estimates) that does not always provide a clear answer to the question that the analysis was conducted to answer. This is particularly true when the regression model includes interaction terms and categorical independent variables. In such cases, the most direct answer is often given by tables of marginal effects. For example, as alluded to in the previous chapter, to compare the average rating of the teacher's actions in the first segment of the breach-breach version of each storyboard made by experienced geometry teachers to the average rating of the teacher's actions in the first segment of the compliance-compliance or compliance-breach version of that storyboard made by such participants, I used a model in which a participant's rating of the teacher's actions in that segment of a given storyboard is regressed on indicators of the condition and storyline of that storyboard and of whether the participant had three or more years of experience teaching geometry, as well as all of the two- and three-way interactions between those three variables. As a reminder, this model (in fact, all of the mixed-effects regression models that I refer to in this chapter) is represented by equation 3.1. In this case, the outcome variable—Y_{ij}—is an indicator of the participant's rating of the teacher's actions in the first segment of a storyboard. To compare an average experience geometry teacher's rating of the teacher's actions in the first segment of the breach-breach version of, for example, the isosceles-triangle storyboard to an average experience geometry teacher's rating of the teacher's actions in the first segment of, for example, the compliance-breach version of the same questionnaire from a table of regression coefficients, one would need to sum β_{0j}, β_{2j}, β_{4j}, β_{6j}, β_{10j}, β_{14j}, β_{16j}, and β_{21j}, sum β_{0j}, β_{1j}, β_{4j}, β_{6j}, β_{9j}, β_{13j}, β_{16j}, and β_{20j}, then take the difference between those two sums. Each of those sums, however, is an example of a marginal effect and can be calculated by the model.

Therefore, rather than present the results of each regression analysis in a regression table, I present them in a 4x3 table of marginal effects in which each cell presents the estimated average experienced geometry teacher score associated with one of the twelve storyboards (e.g., the estimated rating of the teacher's actions in the first segment of that storyboard by an experienced geometry teacher), along with its 95% confidence interval.

4.1. The GCA-Figure Norm

This section presents the results related to the GCA-Figure norm. As a reminder, in attempting to confirm the existence of the GCA-Figure norm, in section 4.1.1, I draw on remarks that the GCA-Figure norm was breached or followed elicited by the first item in each questionnaire (290xy_1). I also draw on remarks that the GCA-Figure norm was breached or followed elicited by the open-ended item in the DRN-style itemsets included in each questionnaire (290xy_3R). In section 4.1.2, to explore experienced geometry teachers' attitudes towards decisions to breach the GCA-Figure norm, I draw on appraisals of the decisions to breach or follow the GCA-Figure norm that were elicited by any item associated with each storyboard. I also draw on responses to the two closed-ended items associated with the GCA-Figure norm: the rating of the teacher's action in the first segment of each storyboard (290xy_2) and the closed-response item in the DRN-style itemsets (290xy_3). I begin with the question of whether the GCA-Figure norm exists.

4.1.1. Does the GCA-Figure Norm Exist?

Table 4.1 presents the proportion of experienced geometry teachers who remarked that the GCA-Figure norm was breached when responding to the first item in the breach-breach version of each storyboard and the proportion of experienced geometry teachers who remarked that the GCA-Figure norm was followed when responding to the same item in the compliance-

compliance and compliance-breach versions of that storyboard (i.e., whose response to that item received a R-GCAF_1 score of 1). In terms of assessing whether the GCA-Figure norm exists, the main take-away from these descriptive statistics is that some (though few) experienced geometry teachers remarked when it was breached, while none remarked when it was followed. Both of these trends are evidence in support of the hypothesis that the GCA-Figure norm exists as normative behaviour tends to go unnoticed while non-normative behaviour tends to attract attention. However, as mentioned above, to determine whether the differences in the proportions of experienced geometry teachers that remarked that the GCA-Figure norm was breached when responding to the breach-breach version of a given storyboard and the proportions of experienced geometry teachers that remarked that it was followed when responding to the compliance-compliance or compliance-breach version of that storyboard are reflective of differences in how any experienced geometry teacher might respond to those storyboards, I compared those proportions using the type of mixed-effects regression model described by equation 3.1. The outcome variable in this case is the score indicating whether a participant

Table 4.1: Proportions of experienced geometry teachers that remarked the GCA-Figure norm when responding to the first item in each questionnaire

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=78	0.00 (0.00) N=106	0.13 (0.34) N=121
	Trapezoid	0.00 (0.00) N=121	0.00 (0.00) N=78	0.04 (0.19) N=105
	Isosceles-triangle	0.00 (0.00) N=78	0.00 (0.00) N=106	0.02 (0.13) N=121
	Parallelogram	0.00 (0.00) N=121	0.00 (0.00) N=76	0.07 (0.25) N=104

remarked that the GCA-Figure norm was followed or breached when responding to the first item associated with a given storyboard (R_GCAF_1). However, given that no one remarked that the GCA-Figure norm was followed when responding to any of the compliance-compliance or compliance-breach storyboards, that model could not estimate the probabilities that an experienced geometry teacher would remark that it was followed when responding to any of the storyboards. Therefore, one approach to determining whether the predicted probability that an experienced geometry teacher would remark that the GCA-Figure norm was breached when responding to the breach-breach version of each storyboard is significantly greater than the predicted probability that an experienced geometry teacher would remark that it was followed when responding to the compliance-compliance or compliance-breach version of that storyboard would be to compare each of the former to 0.00 (the proportion of experienced geometry teachers who remarked that it was followed in the compliance-compliance or compliance-breach versions of that storyboard). However, a more-conservative approach would be to use the *Rule of Three*[73] to estimate 95% confidence intervals of the probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was followed when responding to the compliance-compliance or compliance-breach version of each storyboard, compare each of them to the 95% confidence interval of the predicted probability that an experienced geometry teacher would remark that it was breached when responding to the breach-breach version of that storyboard (calculated by the model), and conclude that the latter is significantly greater than the former if there is no intersection between those confidence intervals. Those confidence intervals are reported in Table 4.2. To the left of each of them, I also report the predicted probability (i.e.,

[73] The Rule of Three states "that 3/n is an upper 95% confidence bound for binomial probability p when in n independent trials no events occur" (Jovanovic & Levy, 1997). In my case, the "independent trials" are participants' responses to a given storyboard and "no event occur[ing]" means everyone receiving a value of 0 with respect to the score being compared in this analysis (R_GCAF_1).

marginal effect) when it could be calculated or the corresponding proportion (reported in Table 4.1) when it could not (i.e., 0.00).

Table 4.2: Predicted probabilities that an experienced geometry teacher would remark the GCA-Figure norm when responding to the first item in each questionnaire

		CONDITION		
		Compliance-breach	Breach-breach	Difference
STORYLINE	Similar-triangles	0.00 (0.00, $\frac{1}{78} \approx 0.01$)	0.00 (0.00, $\frac{1}{106} \approx 0.01$)	0.13 (0.07, 0.19)
	Trapezoid	0.00 (0.00, $\frac{1}{121} \approx 0.01$)	0.00 (0.00, $\frac{1}{78} \approx 0.01$)	0.04 (0, 0.07)
	Isosceles-triangle	0.00 (0.00, $\frac{1}{78} \approx 0.01$)	0.00 (0.00, $\frac{1}{106} \approx 0.01$)	0.02 (-0.01, 0.04)
	Parallelogram	0.00 (0.00, $\frac{1}{121} \approx 0.01$)	0.00 (0.00, $\frac{1}{76} \approx 0.01$)	0.07 (0.02, 0.12)

Based on this analysis, the probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was breached when responding to the breach-breach version of the similar triangles and parallelogram storyboards are significantly higher than the probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was followed when responding to the compliance-compliance and compliance-breach version of those storyboards. These findings support the hypothesis that experienced geometry teachers expect solutions to GCA problems to be geometrically meaningful (i.e., that the GCA-Figure norm exists). While it is possible that this was not the case with the probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was breached when responding to the breach-breach version of the trapezoid or isosceles-triangle storyline is evidence of the contrary, there are other explanations. For example, the fact that so few participants remarked the breach of the GCA-Figure norm in the breach-breach version of the isosceles-triangle storyboard could be because the breach was quite subtle. As a reminder, while the evidence that the GCA-Figure norm is breach in the breach-breach version of the other three storyboards is that the student's calculations result in at least one non-positive side length, the evidence it was breached in the breach-breach version of the isosceles-triangle storyboard is that the calculations result in three positive side lengths that however violate the triangle inequality[74] (13, 13, and 27)—something that a participant would have to check and likely would not if this type of breach is rare in actual classrooms[75]. In regards to the experienced geometry teachers' responses to the breach-breach version of the trapezoid storyboard, it is possible that so few of them remarked that the GCA-Figure norm was breached because experienced geometry teachers actually expect that, when solving GCA problems, the type of figure determined through one's calculations might not be the type of figure represented by the pictorial aspects of a diagram.[76] That said, the next analysis provides some reason to doubt that and instead consider that this might have instead been evidence of an issue with the approach to measure norms of mathematical problems by asking participants to respond to storyboards in which such problems are solved in classrooms, given that participants might not consider the problem something that they are being asked to comment on.

As mentioned in the introduction to section 4.1, another way of answering the question of whether the GCA-Figure norm exists (i.e., research question 1a) is to compare the predicted probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was breached or followed when explaining their rating of the relative appropriateness of the two GCA problems that they were asked to compare in each DRN-style itemsets (i.e., when

[74] As a reminder, the triangle inequality states that the sum of the lengths of any two sides of a triangle is less than the length of the third side.

[75] I did not have the teacher or students in any storyboard remark that the GCA-Figure norm was breached (that the triangle inequality was violated or that a side length was zero of negative) so that I could claim that, when a participant did, it was because they were surprised rather than just being comprehensive in what their description of the storyboard.

[76] In this case, that they could discover that there a trapezoid that meets all the conditions imposed by the algebraic expressions, but there is a triangle that does (if we take 8-2x to be the length of a vertex—0).

responding to item 290xy_3R), across questionnaires. As a reminder, the intended design of the DRN-style itemsets was for the participant to compare the GCA problem in the storyboard to another GCA problem that follows the GCA-Figure norm. As such, this pair of items would ask participants to compare two GCA problems that follow the GCA-Figure norm, when responding to each compliance-compliance or compliance-breach questionnaire, and to compare a GCA problem that breaches the GCA-Figure norm to one that follows it, when responding to the breach-breach version of each questionnaire. However, as mentioned in the previous chapter, there was an error in version of these itemsets included in the trapezoid questionnaires, namely, that the comparison problem (i.e., the one not solved in the storyboard) breaches (rather than follows) the GCA-Figure norm. As such, responses to this itemset in the three trapezoid questionnaires is not comparable and will not be considered in analyses of responses to the DRN-style itemsets (290xy_3 or 290xy_3R).

Table 4.3 presents the proportions of experienced geometry teachers that remarked that the GCA-Figure norm was breached when responding to that item in the breach-breach version of the similar-triangles, isosceles-triangle, and parallelogram questionnaires as well as the proportions of experienced geometry teachers that remarked that it was followed when responding to that item in the compliance-compliance and compliance-breach versions of those questionnaires. Again, what is noteworthy here is that the proportions of experienced geometry teachers that remarked that the GCA-Figure norm was followed when responding to that item in all three compliance-compliance or compliance-breach questionnaires are very low (0.00-0.03), while the proportions of experienced geometry teachers that remarked that the GCA-Figure norm was breached when responding to that item in the breach-breach version of the similar-triangles and parallelogram questionnaires are quite high (0.34). The fact that the proportion of experienced geometry teachers that remarked that the GCA-Figure norm was breached responding to that item in the breach-breach version of the isosceles-triangle questionnaire was lower (0.10)—although still higher than the proportions of participants that remarked that the GCA-Figure norm was followed—could again be due to the breach of the norm being more subtle in the problem presented in the breach-breach version of the isosceles-triangle storyboard (and, therefore, DRN-style itemset).

Table 4.3: Proportions of experienced geometry teachers that remarked the GCA-Figure norm when responding to open-ended item in the DRN-style itemset in each of the similar-triangles, isosceles-triangle, and parallelogram questionnaires

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.01 (0.12) N=69	0.00 (0.00) N=97	0.34 (0.48) N=111
	Isosceles-triangle	0.03 (0.17) N=69	0.00 (0.00) N=100	0.10 (0.31) N=116
	Parallelogram	0.01 (0.09) N=119	0.00 (0.00) N=75	0.34 (0.48) N=99

Towards determining whether these proportions are significantly different, Table 4.4 presents the predicted probabilities that an experienced geometry teacher would remark that the GCA-Figure norm was breached when responding to that item (290xy_3R) in the breach-breach version of each of those questionnaires and the predicted probabilities that an experienced geometry teacher would remark that it was followed when responding to this item in the compliance-breach and compliance-compliance versions of each of those questionnaires. Since the proportion of experienced geometry teachers that remarked that the GCA-Figure norm was followed when responding to that item in the compliance-breach version of each questionnaire was 0.00, I again used the Rule of Three to estimate the 95% confidence intervals of the predicted probabilities that a participant would remark that the GCA-Figure norm was followed by either GCA problem when responding to those items.

Table 4.4: Predicted probabilities that an experienced geometry teacher would remark the GCA-Figure norm when responding to the DRN-style item in each of the similar-triangles, isosceles-triangle, and parallelogram questionnaires

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.01 (-0.01, 0.04)	0.00 (0.00, $\frac{1}{97} \approx 0.01$)	0.34 (0.25, 0.43)
	Isosceles-triangle	0.03 (-0.01, 0.07)	0.00 (0.00, $\frac{1}{100} \approx 0.01$)	0.10 (0.05, 0.16)
	Parallelogram	0.01 (-0.01, 0.03)	0.00 (0.00, $\frac{1}{75} \approx 0.01$)	0.34 (0.25, 0.44)

Based on this, the predicted probability that an experienced geometry teacher would remark that the GCA-Figure norm is breached when responding to that item in the breach-breach version of each questionnaire is significantly higher than the predicted probability that an experienced geometry teacher would remark that the GCA-Figure norm is followed when responding to that item in the compliance-breach version of that questionnaire. Similarly, the predicted probability that an experienced geometry teacher would remark that the GCA-Figure norm is breached when responding to that item in the breach-breach version of the similar-triangles and parallelogram questionnaires is significantly higher than the predicted probability that an experienced geometry teacher would remark that the GCA-Figure norm is followed when responding to that item in the compliance-compliance version of those storyboards[77]. The fact

[77] In the case of similar-triangles questionnaires, the difference is 0.33, SE= 0.05 (Z=6.97, p<0.001). In the case of isosceles-triangles questionnaires, the difference is 0.08, SE= 0.03 (Z=2.20, p<0.05). In the case of parallelogram questionnaires, the difference is 0.33, SE= 0.05 (Z=6.91, p<0.001).

that this was not true of the responses to those versions of the isosceles-triangles questionnaire could again be explained by the breach of the GCA-Figure norm in the breach-breach version of the isosceles-triangle storyboard being subtle. Combined with the very low proportion of experienced geometry teachers that remarked that the norm was followed in response to the compliance-compliance versions of the questionnaires, this is good evidence that the GCA-Figure norm exists. In fact, I argue that this evidence should be prioritized over the evidence from the responses to the first items in each questionnaire. The basis of this argument is that items that ask the participant to evaluate GCA problems are arguably better measures of a teacher's recognition of the GCA-Figure norm than items that ask the participant to evaluate a teacher's actions during a lesson in which a class solves a problem that either breaches or follows it. The basis of that claim is that the typical context in which a teacher evaluates mathematical problems is when they are reading instructional materials (e.g., textbooks) in which problems are presented without distractions, such as how students are thinking about solving it, a teacher is supporting students' work on the problem, etc. Next, I turn to the question of whether experienced geometry teachers deem breaches of the GCA-Figure norm acceptable.

4.1.2. Do Experienced Geometry Teachers Deem Breaches of the GCA-Figure Norm Acceptable?

As mentioned in the introduction to section 4.1, evidence of participants' attitudes towards breaches of either norm comes in two forms: appraisal scores and ratings of the actions taken by the teacher in each storyboard. In each case, these are again compared across the compliance-breach and breach-breach version of each questionnaire (as the storyboards in those pairs of questionnaires differ only in terms of whether the GCA-Figure norm is breached). I begin with the comparison of the positive appraisal score related to the GCA-Figure norm

(PA_GCAF_any). Table 4.5 presents the proportions of experienced geometry teachers that positively appraised the fact that the GCA-Figure norm was followed when responding to at least one item in each of the compliance-compliance and compliance-breach storyboards (i.e., any of the items other than the DRN-style itemset), along with the proportions of experienced geometry teachers that positively appraised the fact that the GCA-Figure norm was breached when responding to at least one item in each of the breach-breach storyboards.

The main take-away from this table is that, similar to how no experienced geometry teachers remarked the decision to follow the GCA-Figure norm in any of the compliance-compliance or compliance-breach storyboards (see Table 4.1), none of them positively appraised those decisions either.[78] Consequently, to statistically compare the proportions of experienced geometry teachers that positively appraised the fact that the GCA-Figure norm was followed when responding to at least one item associated with each of the compliance-compliance and

Table 4.5: Proportions of experienced geometry teachers that positively appraised the decision to follow or breach the GCA-Figure norm when responding to any item associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=69	0.00 (0.00) N=91	0.01 (0.10) N=105
	Trapezoid	0.00 (0.00) N=113	0.00 (0.00) N=67	0.05 (0.21) N=64
	Isosceles-triangle	0.00 (0.00) N=67	0.00 (0.00) N=89	0.02 (0.13) N=113
	Parallelogram	0.00 (0.00) N=107	0.00 (0.00) N=66	0.00 (0.00) N=90

[78] In fact, this could be concluded directly from Table 4.1, as a participant would have had to have recognized that the norm was followed to appraise that decision.

compliance-breach storyboards, along with the proportions of experienced geometry teachers that positively appraised the fact that the GCA-Figure norm was breached when responding to at least one item associated with each of the breach-breach storyboards, I again used the Rule of Three to estimate the 95% confidence intervals of the probabilities that an experienced geometry teacher would positively appraise the fact that the GCA-Figure norm was followed when responding to at least one item associated with each of the compliance-compliance and compliance-breach storyboards. Table 4.6 presents those confidence intervals, as well as the probabilities and confidence intervals that the model could estimate.

Table 4.6: Predicted probabilities that an experienced geometry teacher would positively appraise the decision to follow or breach the GCA-Figure norm when responding to any item associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00, $\frac{1}{69} \approx 0.01$)	0.00 (0.00, $\frac{1}{91} \approx 0.01$)	0.01 (-0.01, 0.03)
	Trapezoid	0.00 (0.00, $\frac{1}{113} \approx 0.01$)	0.00 (0.00, $\frac{1}{67} \approx 0.01$)	0.05 (0.00, 0.06)
	Isosceles-triangle	0.00 (0.00, $\frac{1}{67} \approx 0.01$)	0.00 (0.00, $\frac{1}{89} \approx 0.01$)	0.02 (-0.01, 0.04)
	Parallelogram	0.00 (0.00, $\frac{1}{107} \approx 0.01$)	0.00 (0.00, $\frac{1}{66} \approx 0.02$)	0.00 (0.00, $\frac{1}{90} \approx 0.01$)

According to this analysis, the predicted probability that an experienced geometry teacher would positively appraise the fact that the GCA-Figure norm was breached when responding to at least one item associated with a given breach-breach storyboards is not significantly different from the predicted probability an experienced geometry teacher would positively appraise the fact that it was followed when responding to at least one item associated with the compliance-compliance or compliance-breach version of that storyboard. In other words, this analysis does not provide any evidence that experienced geometry teachers deem breaches of the GCA-Figure norm any more (or less) acceptable than decisions to follow it.

Next, we consider the negative appraisals of decisions to breach or follow the GCA-Figure norm. Table 4.7 presents the proportions of participants that negatively appraised the fact that the GCA-Figure norm was followed when responding to at least one item associated with each of the compliance-compliance or compliance-breach storyboards, as well as the proportions of participants that negatively appraised the fact that the GCA-Figure norm was breached when responding to at least one item associated with each of the breach-breach storyboards. The main take-away from this table is that, similar to how no experienced geometry teachers remarked the decision to follow the GCA-Figure norm in any of the compliance-compliance or compliance-breach storyboards, no experienced geometry teachers negatively appraised those decisions either. Consequently, again, I needed to use the Rule of Three to compare the probabilities that an experienced geometry teacher would negatively appraise the fact that the GCA-Figure norm

Table 4.7: Proportions of experienced geometry teachers that negatively appraised the decision to follow or breach the GCA-Figure norm when responding to any item associated with the storyboard in each questionnaire

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=69	0.00 (0.00) N=91	0.14 (0.35) N=107
	Trapezoid	0.00 (0.00) N=113	0.00 (0.00) N=67	0.12 (0.33) N=65
	Isosceles-triangle	0.00 (0.00) N=67	0.00 (0.00) N=89	0.04 (0.19) N=113
	Parallelogram	0.00 (0.00) N=107	0.00 (0.00) N=66	0.15 (0.36) N=92

was followed when responding to at least one item associated with each of the compliance-compliance or compliance-breach storyboards to the predicted probabilities that an experienced geometry teacher would negatively appraise the fact that the GCA-Figure norm was breached when responding to at least one item associated with the breach-breach version of each storyboard. Table 4.8 presents those confidence intervals and predicted probabilities.

Table 4.8: Predicted probabilities that an experienced geometry teacher would negatively appraise the decision to follow or breach the GCA-Figure norm when responding to any item associated with the storyboard contained in each questionnaire

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00, $\frac{1}{69} \approx 0.01$)	0.00 (0.00, $\frac{1}{91} \approx 0.01$)	0.14 (0.07, 0.21)
	Trapezoid	0.00 (0.00, $\frac{1}{113} \approx 0.01$)	0.00 (0.00, $\frac{1}{67} \approx 0.01$)	0.12 (0.04, 0.20)
	Isosceles-triangle	0.00 (0.00, $\frac{1}{67} \approx 0.01$)	0.00 (0.00, $\frac{1}{89} \approx 0.01$)	0.04 (0.00, 0.07)
	Parallelogram	0.00 (0.00, $\frac{1}{107} \approx 0.01$)	0.00 (0.00, $\frac{1}{66} \approx 0.02$)	0.15 (0.08, 0.22)

The main take-away from this analysis is that the predicted probabilities that an experienced geometry teacher would negatively appraise the decision to breach the GCA-Figure norm when responding to the breach-breach versions of the similar-triangles, trapezoid, and parallelogram storyboards are significantly higher than the predicted probabilities that such teachers would negatively appraise the decisions to follow it when responding to the compliance-compliance and compliance-breach versions of those storyboards (evidenced by the confidence intervals not overlapping). The fact that this was not the case with the isosceles-triangle storyboards could again be due to the subtlety of the breach in the breach-breach version of that storyboard. The fact that the difference in the other six cases, however, is consistent with prior research on norms of instructional situations that demonstrated that teachers prefer decisions to follow norms over decisions to breach them (e.g., Dimmel, 2015).

Having coded each response to each breach-breach storyboard for both whether it contains a negative appraisal of the decision to breach the GCA-Figure norm and whether it contains a positive appraisal of that decision also presents one more way to answer the question of whether experienced geometry teachers tend to deem breaches of the GCA-Figure norm acceptable (or unacceptable). That approach consists of comparing the proportions of experienced geometry teachers that positively appraised the decision to breach the GCA-Figure norm when responding to the breach-breach version of each storyboard to the proportion of experienced geometry teachers that negatively appraise each of those decisions. To make that comparison, I conducted a series of tests of proportions. These tests revealed that the proportion of experienced geometry teachers that negatively appraised the breach of the GCA-Figure norm when responding to the breach-breach version of the similar-triangles and parallelogram storyboards was significantly higher than the proportion of experienced geometry teachers that positively appraised it (even after applying the Bonferroni correction for multiple tests[79])[80]. This suggests that experienced geometry teachers are more likely to deem a breach of the GCA-Figure norm acceptable than unacceptable. The fact that this was not the case of the responses to the breach-breach version of the isosceles-triangle storyboard could be evidence that experienced geometry teachers are not more likely to deem this breach of the GCA-Figure norm acceptable than to deem it unacceptable, but this could again be another consequence of the subtlety of the

[79] According to which *p*-values are multiplied by the number of tests with the same outcome (in this case, four) to account for the increased probability of falsely rejecting the null hypothesis (in this case, that the proportion of experienced geometry teachers in the same who negatively appraised a breach of the GCA-Figure norm is equal to the proportion of experienced geometry teachers in the same who positively appraised it) introduce by running multiple independent comparisons.

[80] In the case of similar-triangles questionnaires, the difference was 0.13, SE= 0.03 ($Z=3.60$, $p<0.001$, before applying the Bonferroni correction). In the case of parallelogram questionnaires, the difference was 0.15, SE= 0.04 ($Z=3.85$, $p<0.001$, before applying the Bonferroni correction).

breach of the GCA-Figure norm in that storyboard. Similarly, the fact that this was not the case of the responses to the breach-breach version of the trapezoid storyboard could be evidence that experienced geometry teachers are not more likely to deem this breach of the GCA-Figure norm acceptable than to deem it unacceptable, but could also be a consequence of some experienced geometry teachers liking the idea of the student's calculations suggesting the existence of a figure of a type different from the figure represented by the diagram (in this case, a triangle rather than a trapezoid).

Next, we consider experienced geometry teachers' ratings of the teacher's actions in the first segment of each storyboard. Table 4.9 presents their average rating of the teacher's actions in the first segment of each scenario (i.e., their responses to item 290xy_2). One take-away from this table is that the average experienced geometry teacher's rating of the teacher's actions during the first segment of each storyboard was between 4.36 and 4.70, which is above the middle of the scale (3.5), suggesting that experienced geometry teachers thought the teacher's

Table 4.9: Average ratings of the teacher's actions in the first segment of each storyboard, among the experienced geometry teachers

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.53 (1.30) N=79	4.49 (1.16) N=107	4.66 (1.12) N=121
	Trapezoid	4.52 (1.32) N=121	4.70 (1.03) N=70	4.46 (1.19) N=105
	Isosceles-triangle	4.72 (0.87) N=78	4.18 (1.18) N=106	4.59 (1.21) N=121
	Parallelogram	4.83 (1.13) N=121	4.77 (0.94) N=77	4.43 (1.24) N=105

actions in the first segment of each storyboard were generally acceptable, whether the problem that they assigned breached or followed the GCA-Figure norm. Yet, since there were other aspects of the first segment of each storyboard that might have made it acceptable (e.g., the teacher giving students time to work in pairs or circulating to check on students), it's unclear whether these average ratings suggest that the experience geometry teachers deemed both the decision to breach the GCA-Figure norm and to follow it acceptable. To determine this, I compared the ratings across questionnaires using another mixed-effects model (again of the type represented by equation 3.1). Table 4.10 presents the probabilities predicted by that model as well as their 95% confidence intervals.

Table 4.10: Predictions of an experienced geometry teacher's rating of the teacher's actions in the first segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.53 (4.27, 4.79)	4.49 (4.27, 4.71)	4.66 (4.45, 4.87)
	Trapezoid	4.52 (4.31, 4.73)	4.70 (4.43, 4.97)	4.46 (4.23, 4.67)
	Isosceles-triangle	4.72 (4.46, 4.98)	4.18 (3.95, 4.40)	4.59 (4.38, 4.79)
	Parallelogram	4.83 (4.63, 5.04)	4.77 (4.51, 5.03)	4.43 (4.20, 4.64)

In terms of whether it provides evidence that experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it, the results are mixed. On the one hand, the predicted rating of the teacher's actions in the first segment ofthe breach-breach version of the parallelogram storyboard is significantly lower than the predicted rating of the same segment of the compliance-compliance and compliance-breach version of the same storyboard.[81] However, the opposite is true in the case of the comparison to the compliance-breach version of the isosceles-triangle storyboard: The predicted rating of the teacher's actions in the first segment of the breach-breach version of that storyboard is significantly higher than the predicted rating of the same segment of the compliance-breach version of the same storyboard.[82] Last, the difference is not statistically significant in the cases of the comparisons to the compliance-compliance or compliance-breach versions of the similar-triangles and trapezoid storyboards, or in the case of the comparison to the compliance-compliance version of the isosceles-triangle storyboard. Therefore, based on this analysis, we cannot say whether experienced geometry teachers deem breaches of the GCA-Figure norm acceptable.

Last, I turn to experienced geometry teachers' ratings of the relative appropriateness of the two GCA problems presented to them in each of the DRN-style itemsets (i.e., their responses to item 290xy_3). Table 4.11 presents the average rating of the relative appropriateness of the two problems presented in this DRN-style item included in the similar-triangles, isosceles-triangle, and parallelogram questionnaires, among the experienced geometry teachers. Again, the responses to this DRN-style item included in the trapezoid questionnaires are not considered in this analysis because of the error in the comparison problem in those items. The main take-away from this table is that the mean rating of the relative appropriateness of the two problems presented in the DRN-style item in the breach-breach version of each of the three questionnaires is above the middle of the 6-point scale: 4.07-4.77. Since these items have the participant compare a problem that breaches the GCA-Figure norm to one that follows it, these results

[81] The difference in the case of the comparison to the compliance-compliance storyboard is 0.41, SE= 0.16 (Z=2.66, $p<0.01$). The difference in the case of the comparison to the compliance-breach storyboard is 0.35, SE= 0.17 (Z=2.01, $p<0.05$).

[82] The difference is 0.41, SE= 0.15 (Z=2.65, $p<0.01$).

suggest that experienced geometry teachers tended to prefer GCA problems that follow the GCA-Figure norm to problems that breach it. The fact that the proportion is lower for the isosceles-triangle questionnaire is again to be expected, given the subtlety of the breach of the GCA-Figure norm in the breach-breach version of the isosceles-triangle storyboard.

Table 4.11: Average rating of the relative appropriateness of the two GCA problems in the DRN-style itemset in each of the similar-triangles, isosceles-triangle, and parallelogram questionnaires, among the experienced geometry teachers

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	3.68 (1.34) N=74	3.60 (0.91) N=105	4.77 (1.33) N=118
	Isosceles-triangle	3.08 (1.00) N=77	3.17 (0.85) N=105	4.07 (1.33) N=121
	Parallelogram	3.46 (1.02) N=120	3.61 (1.15) N=77	4.70 (1.33) N=105

Towards understanding whether these observed differences are reflective of differences that we might see if all experienced geometry teachers were to take the questionnaire, Table 4.12 presents the predictions of an average experience teacher’s response to this item in each similar-triangles, isosceles-triangle, and parallelogram questionnaire. The main take-away from this analysis is that the predicted rating of the relative appropriateness of the two problems was always significantly higher when one of the problems breaches the GCA-Figure norm, again suggesting that experienced geometry teachers tend to prefer GCA problems that follow the GCA-Figure norm over problems that breach it.[83] This aligns with the results of the analysis of

[83] In the case of the comparison to the compliance-compliance version of the similar-triangles questionnaire, the predicted difference is 1.10, SE= 0.17 (Z=6.38, $p<0.001$). In the case of the comparison to the compliance-breach version of the similar-triangles questionnaire, the predicted difference is 1.17, SE= 0.16 (Z=7.52, $p<0.001$). In the case of the comparison to the compliance-compliance version of the isosceles-triangles questionnaire, the predicted difference is 1.00, SE= 0.17 (Z=5.88, $p<0.001$). In the case of the comparison to the compliance-breach version of the isosceles-triangles questionnaire, the predicted difference is 0.90, SE= 0.16 (Z=5.81, $p<0.001$). In the case of the comparison to the compliance-compliance version of the

the negative appraisals of the decision to breach or follow the GCA-Figure norm in each storyboard (presented in Table 4.8).

Table 4.12: Predictions of an experienced geometry teacher's rating of the relative appropriateness of the two GCA problems presented in the DRN-style itemset in each of the similar-triangles, isosceles-triangle, and parallelogram questionnaires

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	3.68 (3.41, 3.94)	3.60 (3.38, 3.82)	4.77 (4.57, 4.99)
	Isosceles-triangle	3.08 (2.82, 3.34)	3.17 (2.95, 3.39)	4.07 (3.87, 4.28)
	Parallelogram	3.46 (3.25, 3.67)	3.61 (3.35, 3.87)	4.70 (4.48, 4.93)

Last, it is worth mentioning again (as I did in the methods chapter) that one trend that we observed during the thematic analysis of the responses to these questionnaires was that many participants suggested that such non-normative GCA problems could be valuable, if the question of whether the figure purported to exist by the statement of the problem actually exists was a topic of whole-class discussion. This is relevant to the question of whether participants deem breaches of the GCA-Figure norm acceptable in that it suggests that this may depend on whether and how the breach is discussed (or, more generally, handled).

4.2. The GCA-Theorem norm

This section presents the results related to the GCA-Theorem norm. In attempting to confirm the existence of the GCA-Theorem norm, in section 4.2.1, I draw on experienced geometry teachers' remarks that a student breached or followed the GCA-Theorem norm,[84] or

parallelogram questionnaire, the predicted difference is 1.24, SE= 0.16 (Z=7.99, p<0.01). In the case of the comparison to the compliance-breach version of the parallelogram questionnaire, the predicted difference is 1.09, SE= 0.17 (Z=6.25, p<0.001).

[84] As a reminder, when the student in a given storyboard breaches the norm, they do so by writing the geometric property that warrants the first equation in their solution, after being asked to do so by the teacher. Similarly, one way that the student in each storyboard follows the GCA-Theorem norm is by verbally stating that property, after being asked to justify the

that the teacher asked them to justify the first equation in their solution, elicited by the first item associated with the second segment of each storyboard (290xy_4). As mentioned earlier, the regression model used to compare the recognition score associated with the GCA-Theorem norm (R_GCAT_4) across questionnaires also provides an answer to the question of whether experienced geometry teachers expectation that the GCA-Theorem norm will be followed depends on whether the GCA-Figure norm is followed. In section 4.2.2, to explore experienced geometry teachers' attitudes towards decisions to breach the GCA-Theorem norm, I draw on experienced geometry teachers' appraisals of a student's decision to breach or follow the GCA-Theorem norm, or the teacher's request that they justify the first equation in their solution, elicited by any item associated with each storyboard. I also draw on experienced geometry teachers' ratings of the teacher's action in the second segment of each storyboard (290xy_5). Again, the regression models used to compare each of these appraisal scores (PA_GCAT_4 and NA_GCAT_4) and ratings (290xy_5) across questionnaires also provide answers to the question of whether experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed. These questions are also answered by comparing experienced geometry teachers' ratings of the teacher's actions throughout each scenario (290xy_6) across questionnaires. Last, in section 4.2.3, I call into question what the GCA-Theorem norm is.

4.2.1. Does the GCA-Theorem Norm Exist?

Table 4.13 presents the proportions of experienced geometry teachers that remarked that the student at the board breached the GCA-Theorem norm and/or that the teacher asked them to

---

first equation in their solution by the teacher. They also follow the norm just prior by writing that equation without justification (again, as the norm is hypothesized to be that the student is expected to either not justify that equation or to justify it verbally).

do so, when responding to the first item associated with the second segment of each storyboard (290xy_4). As a reminder, responses to this item are particularly relevant because those actions take place in the second segment of each of the compliance-breach and breach-breach storyboards, making this item a participant's first opportunity to remark on them.

Table 4.13: Proportions of experienced geometry teachers that remarked that the GCA-Theorem norm was breached in response to the first item associated with the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=77	0.11 (0.32) N=106	0.15 (0.36) N=121
	Trapezoid	0.00 (0.00) N=121	0.32 (0.47) N=77	0.15 (0.36) N=74
	Isosceles-triangle	0.00 (0.00) N=77	0.08 (0.27) N=103	0.12 (0.33) N=121
	Parallelogram	0.00 (0.00) N=121	0.28 (0.45) N=76	0.15 (0.36) N=104

Next, consider the proportions of experienced geometry teachers that remarked one of the ways that the student at the board followed the GCA-Theorem norm and/or interpreted the teacher's request for justification of the first equation in their solution as a request for that justification to be stated verbally, when responding to the first item associated with the second segment of each storyboard (290xy_4), presented in Table 4.14. As a reminder, the student at the board follows the GCA-Theorem norm in each of the storyboards—first by not initially justifying the first equation in their solution, then by verbally stating a geometric property that warrants it, when asked to do so by the teacher. For that reason, one way of answering the question of whether the GCA-Theorem norm exists is to compare the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was breached (and/or that the

Table 4.14: Proportions of experienced geometry teachers that remarked that the GCA-Theorem norm was followed in response to the first item associated with the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.06 (0.25) N=77	0.09 (0.29) N=106	0.12 (0.32) N=121
	Trapezoid	0.10 (0.30) N=121	0.13 (0.34) N=77	0.07 (0.25) N=74
	Isosceles-triangle	0.05 (0.22) N=77	0.06 (0.24) N=103	0.07 (0.25) N=121
	Parallelogram	0.05 (0.22) N=121	0.08 (0.27) N=76	0.07 (0.25) N=104

teacher asked the student to breach it) to the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was followed (and/or that interpreted the teacher's request for justification as a request for the student to verbally state the geometric property that warrants the first equation in their solution), when responding to the first item associated with the second segment of each compliance-breach and breach-breach storyboard. To do so, I conducted a set of eight tests of proportions. The results of these tests mainly suggested that differences in the proportions reported in Tables 4.13 and 4.14 do not reflect differences that one would expect to see in the population of experienced geometry teachers. Specifically, while the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was breached (and/or that the teacher asked the student to breach it) when responding to the compliance-breach version of the trapezoid and parallelogram storyboards was significantly higher than the proportion of experienced geometry teachers that remarked that the GCA-Theorem norm was followed (and/or that interpreted the teacher's request for justification as a request for that justification to be stated verbally) when responding to that version of those storyboards,[85] this was not true of any of the other versions of any of the other storyboards.

The question of whether the GCA-Theorem norm exists can also be answered in a way that leverages the experimental design of the present study. To do that, I compared the predicted probabilities that an experienced geometry teacher would remark that the student in the storyboard breached the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the compliance-breach version of each storyboard to the predicted probability that an experienced geometry teacher would remark that the student followed the GCA-Theorem norm (and/or that interpreted the teacher's request for justification as a request for that justification to be stated verbally) when responding to the compliance-compliance version of those storyboards. As a reminder, these two conditions are compared because storyboards that follow the same storylines in each of those conditions differ only in terms of whether the GCA-Theorem norm is breached. The probabilities and confidence intervals predicted by the mixed-effect model used to make those comparisons are presented in Table 4.15. According to this analysis, in line with the results of the previous tests of proportions, the predicted probability that an experienced geometry teacher would remark that the student breached the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the compliance-breach version of the trapezoid and parallelogram storyboards are significantly higher than the predicted probability that an experienced geometry teacher would remark that the student followed the GCA-Theorem norm (and/or interpreted the teacher's request for justification as a request for that justification to be stated verbally) when responding to the compliance-compliance version of

[85] In the case of the compliance-breach version of the trapezoid questionnaire, the difference was 0.19, SE= 0.07 ($Z=2.89$, $p<0.01$, before applying the Bonferroni correction). In the case of compliance-breach version of the parallelogram questionnaire, the difference was 0.20, SE= 0.06 ($Z=3.18$, $p<0.01$, before applying the Bonferroni correction).

those storyboards.[86] However, again, this was not the case with the similar-triangles or isosceles-triangle storyboards.

Table 4.15: Predicted probabilities that an experienced geometry teacher would remark the GCA-Theorem norm when responding to the first item associated with the second segment of each questionnaire

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.06 (0.01, 0.12)	0.11 (0.05, 0.17)	0.15 (0.09, 0.21)
	Trapezoid	0.10 (0.05, 0.15)	0.32 (0.23, 0.43)	0.15 (0.07, 0.23)
	Isosceles-triangle	0.05 (0.00, 0.10)	0.08 (0.02, 0.13)	0.12 (0.07, 0.18)
	Parallelogram	0.05 (0.01, 0.09)	0.28 (0.18, 0.38)	0.15 (0.08, 0.22)

This same analysis also provides an answer to part of the third research question: Are experienced geometry teachers more likely to expect the GCA-Theorem norm to be followed when the GCA-Figure norm is followed than when it is breached? This question is answered by comparing the predicted probability that an experienced geometry teacher would remark that the student breaches the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the breach-breach version of each questionnaire to the predicted probability that they would make such a remark when responding to the compliance-breach version of that questionnaire. According to the mixed-effects model, the difference is again statistically significant in the cases of the trapezoid and parallelogram storyboards[87], but not in the case of the similar-triangles or isosceles-triangle storyboards. Given that the result for the isosceles-triangle questionnaires might be due again to the subtlety of the breach of the GCA-Figure norm,

[86] In the case of the trapezoid questionnaire, the predicted difference is 0.23, SE= 0.06 (Z=3.84, $p<0.001$). In the case of the parallelogram questionnaire, the predicted difference is 0.23, SE= 0.05 (Z=4.20, $p<0.001$).

[87] In the case of the trapezoid questionnaire, the predicted difference is 0.18, SE= 0.07 (Z=2.73, $p<0.001$). In the case of the parallelogram questionnaire, the predicted difference is 0.13, SE= 0.06 (Z=2.06, $p<0.05$).

I suggest that this set of results be taken as evidence that an experienced geometry teacher's expectation that the GCA-Theorem norm will be followed is dependent on the GCA-Figure norm being followed.

Next, I provide answers to the questions of whether experienced geometry teachers deem breaches of the GCA-Theorem norm acceptable and whether that opinion depends on whether the GCA-Figure norm is breached. I begin with the first of these questions.

4.2.2. Do Experienced Geometry Teachers Find Breaches of the GCA-Theorem Norm Acceptable?

To answer this question, I first compare the proportion of experienced geometry teachers that positively appraised the student's decisions to breach the GCA-Theorem norm (and/or the teacher's request for them to do so) to the proportion of experienced geometry teachers that positively appraised the student's decisions to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request for them to verbally state the geometric property that warrants the first equation in their solution). To begin, Table 4.16 presents the proportion of experienced geometry teachers that positively appraised the decision made by the student at the board to breach the GCA-Theorem norm (and/or the teacher's decision to ask them to do so), when responding to any of the items associated with each storyboard. For sake of comparison, Table 4.17 presents the proportions of experienced geometry teachers that positively appraised the decision made by the student at the board to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request for them to verbally justify the first equation in their solution), when responding to any of the items associated with each storyboard.

One way of answering the question of whether experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it is to compare the proportion of experienced geometry teachers that positively appraised the student's decision to breach the GCA-Theorem norm (and/or the teacher's request for them to do so) to the proportion of experienced geometry teachers that positively appraise the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request for the student to verbally justify the first equation in their solution), when responding to the first item associated with the second segment of each compliance-breach and breach-breach storyboard (since the norm is both followed then breached in each of those storyboards). To do so, I conducted another series of tests of proportions, all of which suggested that the observed differences are not statistically significant. This may be interpreted as evidence that experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it. However, given how small these proportions are, this may simply be evidence that the items were not effective in eliciting appraisals of these decisions.

Table 4.16: Proportions of experienced geometry teachers that positively appraised the decision made by the student at the board to breach the GCA-Theorem norm and/or the teacher's decision to ask them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=69	0.02 (0.15) N=90	0.02 (0.14) N=105
	Trapezoid	0.00 (0.00) N=113	0.01 (0.12) N=67	0.03 (0.18) N=62
	Isosceles-triangle	0.00 (0.00) N=67	0.04 (0.18) N=89	0.03 (0.16) N=113
	Parallelogram	0.00 (0.00) N=107	0.05 (0.21) N=66	0.03 (0.18) N=90

Table 4.17: Proportions of experienced geometry teachers that positively appraised the decision made by the student at the board to follow the GCA-Theorem norm and/or what they interpreted as the teacher's request for them to verbally justify the first equation in their solution, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.01 (0.12) N=69	0.02 (0.15) N=91	0.04 (0.19) N=105
	Trapezoid	0.03 (0.16) N=113	0.01 (0.12) N=67	0.02 (0.13) N=61
	Isosceles-triangle	0.00 (0.00) N=67	0.00 (0.00) N=89	0.02 (0.13) N=113
	Parallelogram	0.02 (0.14) N=107	0.03 (0.17) N=66	0.01 (0.11) N=90

Another way this question can be answered is by comparing the proportion of experienced geometry teachers that positively appraised the student's decision to breach the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the compliance-breach version of each storyboard to the proportion of experienced geometry teachers that positively appraised the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request to verbally justify the first equation in their solution) when responding to the compliance-compliance version of that storyboard. The results of the mixed-effect model used to make those comparisons are presented in Table 4.18.

Based on this analysis, none of the predicted probabilities that an experienced geometry teacher would positively appraise the student's decision to breach the GCA-Theorem norm (and/or the teacher's request for them to do so) when responding to the compliance-breach version of a storyboard are not significantly different from the predicted probability that an experienced geometry teacher would positively appraise the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request to verbally justify the first equation in their solution) when responding to the compliance-compliance version of that storyboard. Therefore, this analysis suggests that experienced geometry teachers do not consider decisions to breach the GCA-Theorem norm acceptable or as acceptable as decisions to follow it. However, again, this is not very strong evidence of this, given how few appraisals were made.

Table 4.18: Predicted probabilities that an experienced geometry teacher would positively appraise the decision made by the student at the board to breach the GCA-Theorem norm and/or the teacher's decision to ask them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.01 (-0.01, 0.04)	0.02 (-0.01, 0.05)	0.02 (-0.01, 0.04)
	Trapezoid	0.03 (0.00, 0.05)	0.01 (-0.01, 0.04)	0.03 (-0.01, 0.08)
	Isosceles-triangle	0.00 (0.00, $\frac{1}{67} \approx 0.01$)	0.04 (0.00, 0.07)	0.03 (0.00, 0.05)
	Parallelogram	0.02 (-0.01, 0.04)	0.05 (0.00, 0.10)	0.03 (0.00, 0.07)

In terms of whether experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed, the predicted probability that an experienced geometry teacher would positively appraise the breach of the GCA-Theorem norm in the breach-breach version of each of the storyboards was not significantly different than the predicted probability that an experienced geometry teacher would positively appraise the breach of the GCA-Theorem norm in the compliance-breach version of that storyboard. This analysis, therefore, suggests that their stances towards breaches of the GCA-Theorem norm do not depend on whether the GCA-Figure norm is followed.

Next, I consider whether experienced geometry teachers were more likely to negatively appraise decisions to breach the GCA-Figure norm than to negatively appraise decisions to follow it. To begin, Table 4.19 presents the proportions of experienced geometry teachers that negatively appraised the decision made by the student at the board to breach the GCA-Theorem norm and/or the teacher's decision to ask them to do so, when responding to any of the items associated with each storyboard. For sake of comparison, Table 4.20 presents the proportions of experienced geometry teachers that negatively appraised the decision made by the student at the board to follow the GCA-Theorem norm and/or what they interpreted as the teacher's decision to

Table 4.19: Proportions of experienced geometry teachers that negatively appraised the decision made by the student at the board to breach the GCA-Theorem norm and/or the teacher's decision to ask them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=69	0.03 (0.18) N=90	0.05 (0.21) N=105
	Trapezoid	0.00 (0.00) N=113	0.07 (0.26) N=68	0.06 (0.25) N=62
	Isosceles-triangle	0.00 (0.00) N=67	0.00 (0.00) N=89	0.01 (0.09) N=113
	Parallelogram	0.00 (0.00) N=107	0.01 (0.12) N=67	0.02 (0.15) N=90

Table 4.20: Proportions of experienced geometry teachers that negatively appraised the decision made by the student at the board to follow the GCA-Theorem norm and/or what they interpreted as the teacher's decision to request that the student verbally justify the first equation in their solution, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=68	0.02 (0.15) N=90	0.01 (0.10) N=105
	Trapezoid	0.00 (0.00) N=113	0.01 (0.12) N=68	0.00 (0.00) N=61
	Isosceles-triangle	0.00 (0.00) N=67	0.02 (0.15) N=89	0.01 (0.09) N=113
	Parallelogram	0.00 (0.00) N=107	0.00 (0.00) N=66	0.01 (0.11) N=90

request that the student verbally justify the first equation in their solution, when responding to any of the items associated with each storyboard.

Again, one way of answering the question of whether experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it is to compare the proportion of experienced geometry teachers that negatively appraised the student's decision to breach the GCA-Theorem norm (and/or the teacher's request for them to do so) to the proportion of experienced geometry teachers that negatively appraise the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's decision to request that the student verbally justify the first equation in their solution), when responding to the first item associated with the second segment of the compliance-breach and breach-breach versions of each storyboard. To do so, I conducted another series of tests of proportions, none of the observed differences were statistically significant. This too may be interpreted as evidence that experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it. However, again, given how small these proportions are, this may simply be evidence that the items were not effective in eliciting appraisals of these decisions.

Another way this question can be answered is by comparing the predicted probabilities that an experienced geometry teacher would negatively appraise the student's decision to breach the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the compliance-breach version of each storyboard to the predicted probabilities that an experienced geometry teacher would negatively appraise the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request to verbally justify the first equation in their solution) when responding to the compliance-compliance version of that storyboard. Those probabilities and their confidence intervals, along with confidence intervals created using the

Rule of Three when no one in a given group (e.g., experienced geometry teachers that were assigned the compliance-breach version of the isosceles-triangle questionnaire) negatively appraised the decision to follow or breach the GCA-Theorem norm, are presented in Table 4.21.

Table 4.21: Predicted probabilities that an experienced geometry teacher would negatively appraise the decision made by the student at the board to breach the GCA-Theorem norm and/or the teacher's decision to ask them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00, $\frac{1}{68} \approx 0.01$)	0.03 (0, 0.08)	0.05 (0.01, 0.09)
	Trapezoid	0.00 (0.00, $\frac{1}{113} \approx 0.01$)	0.07 (-0.01, 0.15)	0.06 (0, 0.12)
	Isosceles-triangle	0.00 (0.00 $\frac{1}{67} \approx 0.01$)	0.00 (0.00 $\frac{1}{89} \approx 0.01$)	0.01 (-0.01, 0.03)
	Parallelogram	0.00 (0.00 $\frac{1}{107} \approx 0.01$)	0.01(-0.02, 0.05)	0.02 (-0.01, 0.06)

Based on this analysis, none of the predicted probabilities that an experienced geometry teacher would negatively appraise the student's decision to breach the GCA-Theorem norm (and/or that the teacher asked them to do so) when responding to the compliance-breach version of a storyboard are significantly different from the predicted probabilities that an experienced geometry teacher would negatively appraise the student's decision to follow the GCA-Theorem norm (and/or what they interpreted as the teacher's request to verbally justify the first equation in their solution) when responding to the compliance-compliance version of that storyboard. Therefore, results of this analysis could also be interpreted as evidence that experienced geometry teachers deem decisions to breach the GCA-Theorem norm as acceptable as decisions to follow it. However, again, given how small these proportions are, this may simply be evidence that the items were not effective in eliciting appraisals of these decisions.

In terms of whether experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed, the predicted probability that an experienced geometry teacher would negatively appraise the breach of the GCA-Theorem norm in the breach-breach version of any of the storyboards was also not significantly different than the predicted probability that an experienced geometry teacher would negatively appraise the breach of the GCA-Theorem norm in the compliance-breach version of that storyboard. This analysis therefore also suggests that their stances towards breaches of the GCA-Theorem norm do not depend on whether the GCA-Figure norm is followed, consistent with the analysis of the positive appraisals. With that said, I move to the analysis of the closed-response data related to the GCA-Theorem norm, beginning with the ratings of the teacher's actions during the second segment of each storyboard, presented in Table 4.22.

Table 4.22: Average rating of the teacher's actions during the second segment of each storyboard, among the experienced geometry teachers

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.69 (1.11) N=78	4.34 (1.23) N=106	4.22 (1.29) N=121
	Trapezoid	4.88 (1.37) N=121	4.55 (1.11) N=76	4.50 (1.30) N=105
	Isosceles-triangle	4.65 (0.99) N=78	4.50 (1.26) N=105	4.85 (1.01) N=121
	Parallelogram	5.18 (1.06) N=121	4.92 (0.85) N=77	4.46 (1.31) N=104

One important take-away from this table is that the average rating of the teacher's actions during the second segment of each storyboard was between 4.32 and 5.10, because this means that people thought the instruction in the second segment of each storyboard was generally acceptable, whether the GCA-Theorem norm was followed or breached. Yet, since there were other aspects of the second segment of each storyboard that might have made it acceptable (e.g., the teacher allowing other students to discuss whether they agree or disagree with the solution presented at the board), it's unclear whether their ratings of the teacher's actions in a given storyboard reflect that they deemed the decision to breach the GCA-Theorem norm and/or the decision to follow it acceptable. To conclude that, I again used a mixed-effects model to compare experienced geometry teachers' ratings of the teacher's actions in the second segment of compliance-breach storyboards to their ratings of the same segment of breach-breach storyboards. Table 4.23 presents those model predictions and their confidence intervals.

In terms of whether experienced geometry teachers deem decisions to breach the GCA-Theorem norm acceptable, the results were mixed. While the predicted ratings of the teacher's actions in the second segment of the compliance-breach versions of the similar-triangles storyboards is significantly lower than the predicted ratings of the teacher's actions in that segment of the compliance-compliance versions of that storyboard[88] and the predicted ratings of the teacher's actions in that segment of the compliance-breach version of the trapezoid storyboard is marginally significantly lower than the predicted ratings of the teacher's actions in that segment of the compliance-compliance version of that storyboard[89], this was not true of the predicted ratings of the teacher's actions in those versions of the isosceles-triangles or parallelogram storyboards. Therefore, it is unclear from this analysis whether experienced geometry teachers consider decisions to breach the GCA-Theorem norm less acceptable than decisions to follow it.

[88] The difference is 0.35, SE=0.17 (Z=2.02, $p<0.05$).
[89] The difference is 0.33, SE=0.17 (Z=1.91, $p=0.06$).

Table 4.23: Predictions of an experienced geometry teacher's rating of the teacher's actions in the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.69 (4.43, 4.95)	4.34 (4.12, 4.56)	4.22 (4.01, 4.43)
	Trapezoid	4.88 (4.67, 5.08)	4.55 (4.29, 4.81)	4.50 (4.28, 4.72)
	Isosceles-triangle	4.65 (4.39, 4.91)	4.50 (4.27, 4.71)	4.85 (4.64, 5.06)
	Parallelogram	5.18 (4.97, 5.39)	4.92 (4.66, 5.18)	4.46 (4.23, 4.68)

In terms of whether experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed, the results were also mixed. While the predicted ratings of the teacher's actions in the second segment of the breach-breach versions of the parallelogram storyboard is significantly lower than the predicted ratings of the teacher's actions in that segment of the compliance-compliance version of that storyboard[90], the opposite was true of the equivalent pair of isosceles-triangle storyboards[91], and there was no significant difference between the equivalent pairs of similar-triangles and trapezoid storyboards. Therefore, this analysis suggests that experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm do not depend on whether the GCA-Figure norm is followed.

A final possible answer to the question of whether experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Figure norm is followed comes from the experienced geometry teachers' ratings of the teacher's facilitation of the work on the problems in the compliance-breach and breach-breach questionnaires. To begin,

[90] The difference is 0.46, SE=0.18 ($Z=2.62$, $p<0.01$).
[91] The difference is 0.36, SE=0.16 ($Z=2.32$, $p<0.05$).

Table 4.24 presents the average rating of the teacher's facilitation of the work on the problem throughout each storyboard among the experienced geometry teachers.

Table 4.24: Average rating of the teacher's facilitation of the work on the problem throughout each storyboard, among the experienced geometry teachers

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.50 (1.07) N=78	4.18 (1.20) N=106	3.66 (1.68) N=121
	Trapezoid	4.36 (1.24) N=121	3.99 (1.29) N=77	3.51 (1.55) N=105
	Isosceles-triangle	4.56 (0,99) N=78	4.09 (1.26) N=105	4.06 (1.53) N=121
	Parallelogram	5.05 (1.04) N=121	5.00 (0.86) N=77	4.06 (1.57) N=105

One thing to take from this table is that the ratings of the teacher's facilitation of the work on the problem throughout each storyboard was between 3.51 and 5.05, suggesting that experienced geometry teachers tended to find the teacher's facilitation of the work on the problem generally acceptable, whether the GCA-Theorem norm was followed or breached. Yet, since there were many other aspects of each storyboard that might have made it acceptable (now that we are referring to the whole storyboard), it's again unclear whether participants' ratings of the teacher's actions reflect that they deemed the decision to breach the GCA-Theorem norm and/or the decision to follow it acceptable. Therefore, again, to conclude that, we use a mixed-effects model to compare the experienced geometry teachers' ratings of the teacher's facilitation of the work on the problem elicited by the breach-breach version of each storyboard to their ratings of the same aspect of the compliance-breach version of that storyboard. Table 4.25 presents those predicted ratings and their confidence intervals.

Table 4.25: Predictions of an experienced geometry teacher's rating of the teacher's facilitation of the work on the problem throughout each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	4.50 (4.21, 4.79)	4.18 (3.93, 4.43)	3.66 (3.42, 3.90)
	Trapezoid	4.36 (4.12, 4.59)	3.99 (3.69, 4.28)	3.51 (3.26, 3.76)
	Isosceles-triangle	4.56 (4.27, 4.86)	4.09 (3.83, 4.33)	4.06 (3.82, 4.29)
	Parallelogram	5.05 (4.81, 5.29)	5.00 (4.70, 5.29)	4.06 (3.80, 4.30)

The main take-aways from this analysis are that the predicted ratings of the teacher's facilitation of the work on the problem in the breach-breach versions of the similar-triangles and trapezoid storyboards are marginally significantly lower than the predicted ratings of the same aspect of the compliance-compliance version of those storyboards[92] and that the predicted ratings of the teacher's facilitation of the work on the problem in the breach-breach versions of the isosceles-triangle storyboard is significantly lower than the predicted ratings of the same aspect of the compliance-compliance version of that storyboard[93]. This was not, however, the case with the parallelogram storyboard, where the differences in the ratings of the teacher's facilitation of the work on the problem in the compliance-breach and compliance-compliance versions were not significantly different. One interpretation of the general trend here is that experienced geometry teachers' attitudes towards breaches of the GCA-Theorem norm depend on whether the GCA-Theorem norm is followed or breached. However, given that the comparison of the ratings of the teacher's actions in the second segment of each storyboard did not provide strong support for this hypothesis and those ratings are arguably better indicators of stances towards breaches of the

[92] In the case of the similar-triangle storyboards, the difference was 0.32, SE=0.20 (Z=1.62, p=0.10). In the case of the trapezoid storyboards, the difference was 0.37, SE=0.19 (Z=1.93, p=0.05).
[93] The difference was 0.48, SE=0.20 (Z=2.45, p<0.05).

GCA-Theorem norm (as they are specific to the second segment of the storyboards in which the norm was followed or breached), I suggest an alternate interpretation: that the differences in the ratings of the teacher's actions throughout the storyboard are actually more indicative of discomfort with multiple norms being breached (i.e., a dosage effect). I come back to this idea in the next chapter.

This concludes the presentation of the results of the analyses that I conducted to determine whether experience geometry teachers are less likely to expect the GCA-Theorem norm to be followed and/or to be more accepting of breaches of that norm, when the GCA-Figure norm is breached, based on my original definition of the GCA-Theorem norm. However, as mentioned in the methods chapter, when coding the open response data, I began to wonder whether I had correctly defined that norm. In the final subsection of this chapter (4.2.3), I consider an alternate definition.

4.2.3. What is the GCA-Theorem norm?

As mentioned in the coding section of the methods chapter (3.3.1), when revising the coding scheme, I separated the code that indicated that a participant remarked that the student at the board followed the GCA-Theorem norm (and/or interpreted the teacher's request for justification of the first equation in their solution as a request for that justification to be stated verbally) into two codes. One of those codes, R_SJ, indicates whether the participant remarked that a student verbally justified that equation and/or interpreted the teacher's request for that justification as a request for it to be stated verbally. The other code, R_NJ, indicates that a student did not justify that equation. I explained that I did this because it was seeming that more participants were remarking the former than the latter, which had me wonder whether the GCA-Theorem norm was actually that the student was only expected to document their algebraic work and that any justification, or request for it, would therefore be unexpected. While the design does not allow for testing that hypothesis by comparing scores across experimental conditions[94], it is possible to compare the proportions of the experienced geometry teachers that remarked that the student at the board justified the first equation in their solution and/or that the teacher asked them to do so to the proportions of the experienced geometry teachers that remarked that the student did not initially justify that equation, when responding to the first item associated with the second segment of each storyboard (290xy_4). Here, I considered any remark that such justification was provided or requested—that is, regardless of whether the participant specified the modality of the provided or expected justification, or not. To begin, Table 4.26 presents the proportions of experienced geometry teachers that remarked that the student justified the first equation in their solution and/or that the teacher asked them to do so when responding to each storyboard.

Table 4.26: Proportions of experienced geometry teachers that remarked that the student justified the first equation in their solution and/or that the teacher asked them to do so elicited by the first item associated with the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.23 (0.43) N=77	0.38 (0.49) N=106	0.41 (0.49) N=121
	Trapezoid	0.31 (0.47) N=121	0.49 (0.50) N=77	0.43 (0.50) N=74
	Isosceles-triangle	0.26 (0.44) N=77	0.28 (0.45) N=103	0.39 (0.49) N=121
	Parallelogram	0.13 (0.34) N=121	0.46 (0.50) N=76	0.42 (0.50) N=104

[94] One limitation of the design of the present study—a consequence of not designing an experimental condition in which the first equation in the solution produced by the student at the board is never justified—that I discuss in the discussion and conclusion chapter.

Table 4.27 presents the proportions of experienced geometry teachers that remarked that the student did not initially justify that equation, when responding to the first item associated with the second segment of each storyboard. This pair of tables demonstrates the trend that the second coder and I noticed: that it was much more common for experienced geometry teachers to remark that the student justified that equation and/or the teacher's request for that justification than to remark that the student did not initially justify that equation.

Table 4.27: Proportions of experienced geometry teachers that remarked that no justification was provided elicited by the first item associated with the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=77	0.05 (0.21) N=106	0.00 (0.00) N=121
	Trapezoid	0.00 (0.00) N=121	0.01 (0.11) N=77	0.01 (0.12) N=74
	Isosceles-triangle	0.00 (0.00) N=77	0.04 (0.19) N=103	0.00 (0.00) N=121
	Parallelogram	0.00 (0.00) N=121	0.04 (0.20) N=76	0.05 (0.21) N=104

To evaluate whether these observed differences are statistically significant, I conducted a series of tests of proportions. In each case, the result was the same: the proportion of experienced geometry teachers that remarked that the student at the board justified the first equation in their solution and/or that the teacher asked them to do so was significantly higher than the proportion of experienced geometry teachers that remarked that the student did not initially justify that equation, when responding to the first item associated with the second segment of each storyboard (290xy_4)[95]. Based on this, it seems that the GCA-Theorem norm may actually be that, within the instructional situation of GCA, students are expected not to justify the first equation in their solutions, leaving the underlying geometric properties (at best) implied[96].

Something else worth stating here is that, while the design of the present study did not allow me to answer the question of what the GCA-Theorem norm is by comparing scores across experimental conditions, it does allow me to provide an answer to the question of whether experienced geometry teachers' expectations that this alternate version of the GCA-Theorem norm will be followed are dependent on whether the GCA-Figure norm is followed. This answer is provided by using the same type of mixed-effects model that I have been using throughout this chapter to compare the predicted probability that an experienced geometry teacher would remark that the student justified the first equation in their solution and/or that the teacher asked them to do so across the breach-breach and compliance-breach versions of each questionnaire. Table 4.28 presents its estimated probabilities and their confidence intervals. According to this analysis, the predicted probabilities that such a comment is made are not significantly different when the GCA-Figure norm is followed than when it is breached, suggesting that experienced geometry teachers' expectations that students will only explain or document their algebraic work when solving GCA problems do not depend on whether the GCA-Figure norm is followed. This result is consistent with findings based on the original definition of the GCA-Theorem norm (presented in Table 4.15).

[95] These differences range from 0.13 to 0.48. The corresponding z-value range from 4.14 to 6.86. The *p*-values are all less than 0.001 (so would be less than 0.01 even after applying the Bonferroni correction for running twelve simultaneous tests).

[96] To be clear, this does not imply that a teacher would not be happy to see a student justify the first equation in their solution, simply that they would be surprised by that decision. Similarly, if the GCA-Theorem norm is instead what I originally hypothesized—that students are expected not to document such justification, this would not imply that they would negatively sanction breaches of that norm. As Bicchieri (2017) explains, breaches of norms may be positively sanctioned. And, as I mentioned earlier, there is an important difference between descriptive and injunctive norms (Cialdini, Kallgren, & Reno, 1991). Importantly, I imagine that the GCA-Theorem norm is an example of the former, as it is hard to imagine a teacher being upset by a student volunteering additional information about their thinking or documentation of it for the class to consider.

Table 4.28: Predicted probabilities that an experienced geometry teacher would remark that the student justified the first equation in their solution and/or that the teacher asked them to do so when responding to the first item associated with the second segment of each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.23 (0.14, 0.32)	0.38 (0.29, 0.47)	0.41 (0.33, 0.50)
	Trapezoid	0.32 (0.23, 0.40)	0.49 (0.38, 0.61)	0.43 (0.32, 0.54)
	Isosceles-triangle	0.26 (0.16, 0.35)	0.28 (0.20, 0.37)	0.39 (0.30, 0.47)
	Parallelogram	0.14 (0.07, 0.20)	0.46 (0.35, 0.57)	0.42 (0.33, 0.52)

Last, the experimental design also allows for answering the question of whether experienced geometry teachers' attitudes towards a student's decision to justify the first equation in their solution—only verbally or also in writing—and/or a teacher's request to ask a student to do so depend on whether the GCA-Figure norm is followed, based on experienced geometry teachers' appraisals of any of those decisions. The measures of experienced geometry teachers' attitudes used in this last set of analyses are the scores representing whether the decision made by the student at the board to justify the first equation in their solution (and/or the teacher's request to ask a student to do so) was/were positively or negatively appraised in any of a participant's responses to a given storyboard. I begin with the results related to the positive appraisals. Table 4.29 presents the proportion of experienced geometry teachers that positively appraised the decision made by the student at the board in each storyboard to justify the first equation in their solution (and/or the teacher's request for them to do so), when responding to any of the items associated with each storyboard. Table 4.30 presents the confidence intervals of the predicted probabilities that an experienced geometry teacher would positively appraise the decision made by the student at the board to justify the first equation in their solution (and/or the teacher's request for them to do so), when responding to any of the items associated with each storyboard.

Table 4.29: Proportions of experienced geometry teachers that positively appraised the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.13 (0.34) N=69	0.13 (0.34) N=90	0.12 (0.33) N=105
	Trapezoid	0.10 (0.30) N=113	0.06 (0.24) N=67	0.13 (0.34) N=63
	Isosceles-triangle	0.04 (0.21) N=67	0.09 (0.29) N=89	0.10 (0.30) N=113
	Parallelogram	0.06 (0.25) N=108	0.14 (0.35) N=66	0.08 (0.27) N=90

Table 4.30: Predicted probabilities that an experienced geometry teacher would positively appraise the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.13 (0.05, 0.21)	0.13 (0.06, 0.20)	0.12 (0.06, 0.18)
	Trapezoid	0.10 (0.04, 0.15)	0.06 (0.00, 0.12)	0.13 (0.04, 0.19)
	Isosceles-triangle	0.04 (0.00, 0.09)	0.09 (0.03, 0.15)	0.10 (0.04, 0.15)
	Parallelogram	0.06 (0.02, 0.11)	0.14 (0.05, 0.22)	0.08 (0.02, 0.13)

As demonstrated by Table 4.29, a few participants positively appraised the student's decision to justify the first equation in their solution and/or the teacher's request for justification in each storyboard. However, given how low the proportions are, this data should not be taken as evidence that experienced geometry teachers are in favour (or against) such decisions. Moreover, according to the regression analysis, the probability that an experienced geometry teacher would positively appraise the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for justification, when responding to a breach-breach storyboard is not significantly different than when responding to a compliance-breach storyboard. This suggests that experienced geometry teachers' attitudes towards a student's decision to share the geometric reasoning behind their algebraic work or a teacher's request for that justification are not dependent on whether the problem follows the GCA-Figure norm.

Last, it is worth similarly considering the negative appraisals of the student's decision to justify the first equation in their solution and/or the teacher's request for them to do so. Table 4.31 presents the proportion of experienced geometry teachers that negatively appraised the student's decision to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to any of the items associated with each storyboard. Table 4.32 presents the confidence intervals of the predicted probabilities that an experienced geometry teacher would negatively appraise the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to

Table 4.31: Proportions of experienced geometry teachers that negatively appraised the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0.00) N=68	0.04 (0.21) N=90	0.05 (0.21) N=105
	Trapezoid	0.00 (0.00) N=113	0.07 (0.26) N=68	0.08 (0.27) N=62
	Isosceles-triangle	0.00 (0.00) N=67	0.02 (0.15) N=90	0.01 (0.09) N=113
	Parallelogram	0.00 (0.00) N=107	0.04 (0.21) N=67	0.03 (0.18) N=90

Table 4.32: Predicted probabilities that an experienced geometry mathematics teacher would negatively appraise the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for them to do so, when responding to any of the items associated with each storyboard

		CONDITION		
		Compliance-compliance	Compliance-breach	Breach-breach
STORYLINE	Similar-triangles	0.00 (0, $\frac{1}{68} \approx 0.01$)	0.04 (0.00, 0.09)	0.05 (0.01, 0.09)
	Trapezoid	0.00 (0, $\frac{1}{113} \approx 0.01$)	0.07 (0.01, 0.15)	0.08 (0.01, 0.14)
	Isosceles-triangle	0.00 (0, $\frac{1}{67} \approx 0.01$)	0.02 (-0.01, 0.05)	0.01 (-0.01, 0.03)
	Parallelogram	0.00 (0, $\frac{1}{107} \approx 0.01$)	0.05 (0.00, 0.10)	0.03 (0.00, 0.08)

any of the items associated with each storyboard.

As demonstrated by Table 4.31, no experienced geometry teachers negatively appraised the student's decision to verbally justify the first equation in their solution, after being asked by the teacher to justify that equation, when responding to any of the compliance-compliance storyboards. In contrast, a few experienced geometry teachers negatively appraised the decision to justify that equation and/or the decision to ask for that justification when responding to the compliance-breach and breach-breach storyboards, in which the justification was both stated verbally and written. However, according to the regression analysis, the probability that an experienced geometry teacher would negatively appraise the decision made by the student at the board to justify the first equation in their solution and/or the teacher's request for justification, when responding to a breach-breach storyboard is not significantly different than when responding to a compliance-breach storyboard. This result also suggests that experienced geometry teachers' attitudes towards a student's decisions to share the geometric reasoning behind their algebraic work or a teacher's request for that justification are not dependent on whether the problem follows the GCA-Figure norm, similar to the analysis of the positive appraisals.

This concludes the presentation of the results of the present study. In the next chapter, I offer a more thorough discussion of these results as well as the limitations of the present study and some potential directions for future research.

Chapter 5 - Discussion and Conclusions

In the introduction to the dissertation, I provided a number of reasons for studying the instructional situation of GCA in U.S. high school geometry. One of the reasons is that GCA is amongst the most common instructional situations in U.S. high school geometry, making it one of the most common opportunities that students have to learn the content of that course. I also argued that the two norms that I hypothesized to exist are worthy of study because, if they exist (i.e., if my statements of those norms each represent an expectation about the instructional situation of GCA that is shared amongst teachers of that course), then they constrain those opportunities to learn. As a reminder, one thing that students could learn through work on GCA tasks is the set of geometric properties that they need to use to solve the GCA problems that they are assigned. The GCA-Theorem norm constrains those opportunities in the sense that, if students are not expected to document those properties, then those properties are less likely to be one of the topics of the whole class discussion of students' proposed solutions to GCA problems. Again, my original hypothesis about the GCA-Theorem norm was that it would be normative for students to verbally state the geometric property that warranted the equation(s) that they set up and solve towards determining dimensions of a geometric figure, and that a teacher could ask them to provide such justification, but that students may also be regularly permitted to only document and talk through their algebraic work. The data provided some evidence that this is the norm, but also some reason to doubt my hypothesis. More specifically, both the tests of proportions and mixed-effects regression analysis that I conducted (reported in section 4.2.1) suggested it was more common for participants to remark that the student at the board wrote down that justification or that the teacher asked them to do so than it was to remark that the student did not initially provide that justification, that the teacher asked them to do so, or that the student then stated it verbally, when responding to storyboards that followed two of the four storylines.[97] However, when I compared the proportion of experienced geometry teachers that mentioned that such justification was provided or requested to the proportion of that remarked that no justification was provided (in section 4.2.3), the former was significantly higher, suggesting that the GCA-Theorem norm might be that students are only expected to document and talk through their algebraic work. Whichever the norm is, however, participants rarely appraised decisions to follow or to breach it, and rated segments of the storyboards in which the teacher insisted that the justification be written similarly to how they rated segments of the storyboards in which the teacher accepted the student's verbal justification as sufficient. One interpretation of this result is that participants are indifferent about whether students communicate—verbally or in writing—the geometric reasoning behind their calculations. Another is that the items included in the surveys were not effective measures of their attitudes towards decisions to follow or breach this norm. I return to this point in the section of this chapter on limitations of the present study.

The other norm that I hypothesized to exist—that I call the GCA-Figure norm—constrains opportunities that students have to learn in the instructional situation of GCA differently. In the introduction, I argued that mathematics education researchers might be interested in the instructional situation of GCA because GCA tasks require students to reason with and across mathematical representations—a mathematical practice that our field has

[97] As a reminder, the INR GCA instrument consists of twelve questionnaires. Each includes a storyboard which belongs to one of three experimental conditions and follows one of four storylines. Multiple storyboards within each experimental condition were created to improve the content validity of any general claims that I would make about the instructional situation of GCA on the basis of comparisons of scores across experimental conditions.

deemed important for decades (e.g., NCTM, 1989, 2000, 2018). Specifically, GCA tasks ask students to calculate dimensions of a geometric figure, initially represented by pictorial aspects of a provided diagram (strokes and spaces) as well as by algebraic expressions, requiring them to set up equations on the basis of geometric properties, solve those equations using algebraic skills learned in Algebra 1 and prior, then use the solution to evaluate those algebraic expressions. Yet this situation could also offer a fairly unique opportunity in U.S. high school mathematics education, as GCA problems can be designed to require students to use the outcome of those calculations to question the existence of the geometric figure purported to exist by the statement of the problem and diagram. Yet, my hypothesis was that GCA problems are not typically designed to engage this conversation. My statement of the GCA-Figure norm represents my hypothesis that experienced geometry teachers expect solutions to GCA problems to evidence the existence of a particular geometric figure that satisfies all the characteristics of the diagram (i.e., a figure of the type represented by the pictorial aspects of the diagrams with dimensions that can be related by the algebraic expressions used to represent them). The analyses presented in section 4.1 provide some evidence that this is an expectation held by experienced geometry teachers, particularly the responses to the DRN-style itemset, which I argued is a better measure of teachers' expectations about GCA problems[98] than items that elicit teachers' reactions to storyboards in which the problem may be seen as part of the background, rather than something to analyze and comment on. Moreover, in terms of experienced geometry teachers' attitudes towards decisions to follow or breach the GCA-Figure norm, the analysis based on the appraisal

[98] In fact, I suggest this would be a better measure of any norm that regulates the types of problems that are assigned in a given instructional situation—a hypothesis that could be tested in future studies.

scores and appropriateness ratings (presented in section 4.1.2) suggests that such teachers may take issue with problems that require students to question the existence of the geometric figure purported to exist by the statement of the problem and diagram. While I would maintain that there is value in breaching or changing this norm, it is worth noting that this aversion is understandable: If it was very uncommon for students to solve GCA problems that require them to question the existence of the figure,[99] it may take students some time to understand the results of their calculations, making such GCA problems less efficient measures of students' understanding of the geometric properties. I imagine their efficiency—in contrast to, for example, the time it takes to compose a geometric proof—is one reason why they are so common. I take this finding and hypothesized reason for the existence of this norm to be amongst the most important results of the present study, as the description of this potential trade-off—between efficiency and prolonged time on task to offer a unique and important opportunity to learn—represents some of the complexity of the work of teaching. Both this and the findings about the GCA-Theorem norm are also useful in that they help us understand a critique of the inclusion of GCA problems in the geometry curriculum: that GCA problems are simply opportunities to rehearse previously-learned algebraic skills (shared with me in casual conversations with both geometry teachers and mathematicians). The fact that breaches of both the GCA-Figure and GCA-Theorem norms foreground geometric reasoning demonstrates why they are valuable: they legitimize the inclusion of GCA problems in the geometry curriculum.

[99] Evidence that it is rare including no participants remarking that it was followed when responding to storyboards in which it was followed and very few participants remarking it when responding to DRN-style itemsets in which both of the presented problems followed.

Finally, one of the answers to the third research question (presented in section 4.2.3) is that experienced geometry teachers' expectations that students might (or might not) share the geometric reasoning behind their calculations appear to be independent of whether the GCA-Figure norm is followed. However, another answer (presented in section 4.2.1) is that they are less likely to expect that reasoning to be included in the student's written work when the GCA-Figure norm is breached. As I suggested in the introduction, one reason this may be is that the breach of the GCA-Figure norm might have had some participants reframe the task as novel (Doyle, 1988) and therefore abandon their expectations that norms of the instructional situation of GCA would be followed. However, this should not be taken as evidence that participants approved of the teacher's insistence that the student document the geometric reasoning behind their algebraic work. To the contrary, the predicted rating of the teacher's facilitation of the work was significantly lower when both the GCA-Figure and GCA-Theorem norms were breached than when only the GCA-Figure norm was breached and the teacher insisted that the student document the geometric reasoning behind their algebraic work.

As mentioned in the introduction, these explorations of whether and how the GCA-Figure norm and GCA-Theorem norm are related, as well as how those relationships can be studied, are arguably the most important contributions of the present study. This is because prior research on norms of mathematics instruction had not investigated potential dependencies between the multiple norms that apply to a particular situation. In the next section, I elaborate on this point.

5.2. Theoretical and Methodological Contributions

The present study makes a number of contributions to both the literature on geometric calculation (GC) and the literature on norms of mathematics instruction. In terms of its contributions to the literature on GC, it demonstrates limitations of studies of GC problems and tasks through providing insights gained by studying the instructional situations in which they exist in classrooms. The present study does this by demonstrating that students' work on GCA problems is regulated by norms, which may not be evident from an analysis of problems in textbooks or triggered (in the mind of the student) when students are asked to solve GCA problems in lab settings. Therefore, to gain a full understanding of the opportunities to learn provided by GCA problems, we need to study how such problems are solved in classrooms (events that are typically framed as instances of the instructional situation of GCA).

In terms of its contribution to the literature on norms of mathematics instruction, a modest one is that it adds to a growing body of evidence that teachers have expectations about the tasks that students are assigned, how they will be solved, and what feedback a teacher will provide, as well as attitudes towards decisions that breach those expectations (literature reviewed in section 2.2.4). While the results of the quantitative analysis aligned with prior research that breaches of norms (in this case, the GCA-Figure norm) are often disapproved (e.g., Garfinkel, 1963), the qualitative analysis demonstrated that teachers might find some breaches quite reasonable (Herbst et al., 2018), in particular under certain conditions. For example, some participants' responses suggested that they would be open to assigning problems that breach the GCA-Figure norm, if the class was engaged in a whole class discussion about what the result of the calculations tells them about the existence of the figure. Similarly, some participants' responses suggested that they thought requests for a student to share the geometric reasoning would be quite reasonable if they came before a student begins solving a GCA problem at the board or after they are done solving it (rather than while they are solving it). The fact that they thought this is a request that the teacher would need to make is also additional evidence that students are not expected to share that geometric reasoning.

As mentioned earlier, the present study also extends our understanding of norms by demonstrating that two norms may be related, for example, in the sense that teachers' expectations that one norm will be followed may depend on whether other norms are followed. In the introduction and above, I provided one reason why this might be the case: that a breach of a norm could cause someone to second-guess their initial framing of the situation. This seems particularly likely if the norm that is breached is *central* (e.g., Bicchieri, 2006) to that frame, in the sense that the observation that it is followed typically has people frame the situation they are in as an instance of the type of situation to which that norm applies. Along those lines, the present study has implications for instructional improvement: While some have suggested that, when policy makers, education researchers, or teacher educators propose changes to mathematics instruction, they need to consider which norms they are (intentionally or unintentionally) asking teachers and students to breach (Herbst, Nachlieli, & Chazan, 2011), my results suggest that they also need to consider which other norms teachers and students may abandon as a consequence. Along those lines, the methodology that I employed—having teachers interact with representations of reform instruction—is useful in that it can be used to predict both which norms a proposed reform will breach and some of the consequences of those breaches.

With that said, I move to discussing some of the limitations of the present study. These include limitations due to the experimental design (the conditions themselves and the assignment of participants to conditions), the design of the storyboards, the design of the items, and the design of the coding scheme.

5.3. Limitations

5.3.1. Limitations Due to the Experimental Design

One limitation of the experimental design of the present study is the way participants were assigned to questionnaires. As mentioned in the methods chapter, I randomly assigned participants to one of three groups and assigned each group to four questionnaires, two in one condition and two in another. This assignment was used as one way to handle the possibility that the three groups, while each made up of a large, random subset of the participants, might not be equal (on average) in terms of some measured and/or unmeasured individual characteristics (e.g., years of experience teaching U.S. high school geometry). If the groups proved to be unequal (in this sense), I imagined that I could conduct within-group-across-condition analyses. However, in hindsight, I realize that this had a couple of negative consequences. First, it would greatly reduce the reliability of the estimates as they would only be based on roughly one third of the sample. Second, even if the groups were equal,[100] this design meant exposing participants to two experimental conditions and accepting that a threat to the validity of the claims that I have made is that participants might remark a decision to follow either norm because they previously saw a storyboard in which it was breached, rather than because their expectation was breached. To handle that, I could have included a fixed effect that represented whether the participant who contributed a given response saw a storyboard that belonged to a different experimental

[100] Of course, this is never entirely the case, either in terms of their level of measured constructs (e.g., amount of relevant knowledge) and levels of unmeasured constructs.

condition prior. Yet, I did not do this because of issues of collinearity with some of the condition-by-storyline interaction terms caused by the fixed order in which each group of participants was assigned each of their four questionnaires. For these reasons, if I were to rerun this experiment, I would randomly assign each participant to only one experimental condition and randomize the order in which the four questionnaires in that condition are presented to each participant.

A second limitation of the study is due to the design of the compliance-breach and breach-breach conditions, in particular, that the GCA-Theorem norm is both followed and breached in each storyboard. As a reminder, in the second segment of all storyboards, the student writes an equation, begins to solve it, then is asked by the teacher to justify the first equation in their solution, then verbally justifies it. All of this was hypothesized to be within the range of what is acceptable according to the GCA-Theorem norm (as I originally proposed it). Then, in the compliance-breach and breach-breach storyboards, the teacher tells the student that they should have written that justification, after which the student does so.[101] As explained in the methods chapter, the rationale behind this design was that the compliance-breach and breach-breach conditions represent how a teacher in a real classroom might ask their students to write the justification if it was normative for them to do so. For sake of comparison, consider the instructional situation of doing proofs (Herbst, 2006) in which students are expected to write the justification for each statement that they make (Herbst, Chen, Weiss, & Gonzalez, 2009). In that situation, if a student forgot to include a reason, a teacher would likely ask them what it is,

[101] Note that this is unlike the difference between the breach-breach storyboards and their paired compliance-breach and compliance-compliance storyboards, which differ in the GCA-Figure norm where it is breached in the former and followed in both of the latter. This design is similar to how, in drug trials, one group receives the drug and another group receives a placebo. In some trials, a group may receive the placebo followed by the drug, but typically a second group receives the drug followed by the placebo, a third group receives the drug twice, and a fourth group receives the placebo twice. The goal in these cases, however, is different: to determine the effect of receiving the drug twice (which was not the case in the present study).

expecting that they would know they need to include it in their written work. However, a student (especially one who has not composed many proofs) might simply reply verbally. In this case, the teacher would likely provide feedback on the justification provided, then ask the students to include it in their written work. While I would maintain that all of the storyboard representations of the instructional situation of GCA in the INR instrument are realistic, having storyboards in which the GCA-Theorem norm was followed then breached is not ideal for testing whether written justification is expected or unexpected in the instructional situation of GCA. Instead, I should have had the teacher in the compliance-breach and breach-breach conditions only ask that they include the justification in their written work. The consequence of not having done so is that it is impossible to know whether a participant commented on the written justification, or the teacher's request for that justification to be written, because they were surprised by either decision (evidence that the GCA-Theorem norm does not include an expectation that such justifications would be written) or because they were surprised by the teacher's insistence that it be stated once it had been written (in cases when participants did not say that in their response to a given item).

A third limitation of the present study is due to the fact that I did not include a breach-compliance condition (in which the GCA-Figure norm would have been breached and the GCA-Theorem norm would have been followed). I did not include this condition because it did not originally seem necessary to answer my research questions. I could (and did) answer questions about the GCA-Figure norm by comparing scores associated with the first segment of each storyboard across the compliance-breach and breach-breach conditions. I could (and did) answer questions about the GCA-Theorem norm by comparing scores associated with the second segment of each storyboard across the compliance-compliance and compliance-breach conditions. Last, I could answer questions about whether expectations that the GCA-Theorem norm would be followed, and/or attitudes towards breaches of it, depend on whether the GCA-Figure norm is followed by comparing scores associated with the segment of each storyboard across those same two conditions (compliance-breach and breach-breach). However, I had not anticipated using the last item in each questionnaire to try to answer the third research question, which is where the breach-compliance condition would come in handy. I included the itemset that asks participants to evaluate the teacher's facilitation of the work on the problem throughout the storyboard in each questionnaire only to capture reactions to the breach of the GCA-Figure norm in a breach-breach storyboard from participants who did not realize that the norm was breached until the problem had been solved at the board. I had not anticipated that it could be used as another measure of whether reactions to the breach of the GCA-Theorem norm depends on whether the GCA-Figure norm is followed. However, my analysis suggests that the predicted ratings of the teacher's facilitation of the work in the breach-breach storyboards were lower than the predicted ratings of the teacher's facilitation of the work in the compliance-breach storyboards, I was left unsure whether the result was evidence of an objection to two norms (rather than one) being breached or to the GCA-Theorem norm being breached when the GCA-Figure norm was also breached. If I had included a breach-compliance condition, I could have answered that question by seeing whether the differences between reactions to storyboards in which both norms were breached and reactions to storyboards in which only the GCA-Theorem norm was breached were similar to the differences between reactions to storyboards in which both norms were breached and reactions to storyboards in which only the GCA-Figure norm was breached. If they were the same, the lower rating of the storyboards in which the two norms were breached would clearly be evidence of aversion to multiple norms being breached. If they were different, it could be that the item was capturing a dependency between the two norms (e.g., that breaches of the GCA-Figure norm have teachers reframe the situation and consequently have different expectations and attitudes towards a student's description of their geometric reasoning). Without a breach-compliance condition, I went with what seemed like the more conservative of the two inferences: objection to two norms being breached.

The final limitation of the experimental design is that it did not allow me to test whether the GCA-Theorem norm is what I originally hypothesized or whether it is that U.S. high school geometry students are expected to document their algebraic work, but not expected to share the geometric reasoning behind it, verbally or in writing. To test that, I could have included a condition in which the student at the board does not justify the first equation in their solution and the teacher does not ask them to do so, at any point in the storyboard. As mentioned in the methods chapter, I did not do this in the current version of the experiment because my interest was in whether their expectations about the modality of the justifications provided by students and/or requested by the teacher. But, given how common it was to remark that a justification was provided and/or requested, and how uncommon it was to remark that a justification was not initially provided, if I were to rerun the experiment, I would include such a condition.

5.3.2. Limitations Due to the Storyboards

One limitation of the present study is due to the fact that the student in the similar-triangles storyboards justifies their choice to equate the algebraic expression that represents the length of segment $\overline{AO}$ in triangle AOC and the algebraic expression that represents the length of segment $\overline{DO}$ in triangle BOD by claiming that $\overline{AO}$ and $\overline{DO}$ are corresponding sides in similar triangles. This was an issue because some participants critiqued this, claiming that the student should have explained how they know that the two triangles are similar, but did not explicitly acknowledge the justification that the student provided. It was therefore unclear whether to code such responses as evidence that the participant did not recognize that a justification was provided or as evidence that they recognized that a justification was provided but objected to the lack of feedback on that justification. I decided to assume that the latter was the case, as it seemed more likely, but that decision introduces potential error to my inferences from the data about whether experienced geometry teachers expect students to justify the first equation in their solutions.

A second issue with the current storyboards is that, in each storyboard, after solving the problem, the student at the board summarizes their work by explaining that they determined their solution by solving the equation that they set up then using the solution to evaluate certain algebraic expressions in order to determine the side lengths that the problem requested. This was an issue because many participants wrote that the student "explained" or "clarified" their work, which may have been reference to the student's justification of the first equation in their solution, but could not be coded as such, because it could be reference to the student's summary of their work. While this felt like the right decision—again trying to be as conservative with my inferences as possible—its consequence is that I missed all of the cases in which this was a participant's attempt to remark that they recognized the justification (and its modality, in the case of "explained"), which is relevant to the analyses that I conducted to determine whether and how experienced geometry teachers expect their students to share their geometric reasoning. Knowing now that people often use vague language to refer to moments in a storyboard, if I were to rerun the experiment, I would not have the student summarize their work or the teacher ask them to do so, so that any reference to the student explaining, elaborating, or clarifying could be attributed to the participant noticing the justification and/or the request for it (depending on the response).

A final potentially productive revision to the experiment would be to make some of the dimensions that the student is asked to determine angle measures, rather than only side lengths. This is common in GCA problems and, therefore, the fact that all of the problems that I created had the student determine side lengths threatens the content validity of my claims about the instructional situation of GCA.

5.3.3. Limitations Due to the Coding Scheme

The coding of the open-response data was a substantial effort. It took months to code the data with the second coder, between the time needed to train them on the original coding scheme, deciding whether and how to revise the coding scheme and codebook, applying the revised coding scheme to a subset of the data with him, and reconciling the differences between our applications of the final version of the coding scheme. As a consequence, as mentioned in the methods chapter, I decided to develop the set of search terms that I used to code the remainder of the data. This also took quite a lot of time, between analyzing all the responses the second coder and I analyzed together to determine a large set of search terms, trying to apply all of them and realizing how long that would take, choosing a subset of terms that were most effective, then using those to code the remainder of the data. The use of the search terms, however, while necessary in the context of completing a dissertation in a (somewhat) timely fashion, introduced some error into the analysis. Namely, since I only coded each response located by a search term, the use of the search terms no doubt had me not apply codes to responses to which they would have been applied had I coded all 11,700 responses. That said, as I mentioned in the methods chapter, I stand by the results as they were based on comparisons of scores (based on those codes) across conditions and I could not think of a reason why any of the terms would have me miss more responses to storyboards in one condition than in another.

A second limitation caused by my choice to be conservative in my application of each code is that I might have missed evidence of appraisals. For example, some (although not many) participants referred to the proposed solution to GCA problems that breached the GCA-Figure norm as an “unreasonable answer” or as “nonsensical”. I chose to only code these as evidence that the participant remarked that the GCA-Figure norm was breached, rather than also as negative appraisals of decisions to breach that norm. While the comparison of the negative appraisal scores across experimental conditions still provided evidence that experienced geometry teachers are more likely to disapprove of a decision to breach the GCA-Figure norm than to follow it, it could be that I underestimated the estimated difference because of this coding decision.

Last, in response to the open-ended item in each of the DRN-style itemsets that had participants compare a problem in which the GCA-Figure norm was breached to one in which it was followed, some participants remarked that one of the two problems followed the GCA-Figure norm by describing that it did not breach it (e.g., “the triangle actually exists”). I coded each of these as recognition that the norm was followed, although it could well have been the case that they only provided that justification for their rating because they recognized the breach. This did not stand in the way of demonstrating that participants remarked problems that breached the GCA-Figure norm more often than they remarked problems that followed it (suggesting that my statement of that norm does represent an expectation held by experienced geometry teachers), but again could have had me underestimate the difference.

5.3.4. Limitations Due to the Items

One issue with the design of the items had to do with the DRN-style itemsets. The issue was that the participant would need to solve each problem in order to realize whether the GCA-

Figure norm was breached or followed in each of the problems. The only exception would be if they realized that one of them was the one being solved in the storyboard or noticed the breach from the student work that they see when viewing the first segment of each storyboard (before answering the DRN-style itemset). But it was clear that at least some participants didn't realize this: some explained that they couldn't see any difference between two problems when only one breached the norm. The consequence of this, again, is that more participants might have remarked that the GCA-Figure norm was breached or followed if they had noticed that it was breached or followed, which might have given a different answer to the question of whether the GCA-Figure norm exists. For that reason, if I were to re-conduct the experiment, I would mention that one of the problems is the one solved in the storyboard and include a solution to each problem. This would also increase the probability that participants who expect the GCA-Figure norm to be followed notice that the problem in the breach-breach storyboards breaches it, before viewing the rest of the storyboard, which would allow me to better measure any potential relationship between the two norms.

Last, I would also revise the set of questions that ask the participant to rate the appropriateness of the teacher's actions. As explained in the literature review and methods chapters, I borrowed these items from earlier INR instruments used to investigate norms about how a teacher should act (e.g., whether they should provide the proposition that students are asked to prove; Herbst, Aaron, Dimmel, & Erickson, 2013). In contrast, the GCA-Figure norm is a rule about GCA problems and, as mentioned earlier, a participant may not consider the problem on the board as one of the things they are being asked to evaluate. In fact, this was explicitly mentioned by one participant in their explanation of their rating of the appropriateness of the teacher's facilitation of the work throughout a storyboard (29003):

> The teacher never addressed the problem of negative side lengths. **If you include problem selection and design as part of the facilitation**, then the teacher's actions were very inappropriate. **However, I don't think that was the point of this scenario** and there were no other glaring weaknesses.

Similarly, the GCA-Theorem norm is a rule about students' solutions and, while the current question could motivate someone to evaluate the teacher's decision to ask the student to justify the first equation in their solution (verbally or in writing), it might not prompt them to evaluate the student's choice to provide that justification. As a consequence, it is possible that these items were less effective than possible alternatives in measuring participants' attitudes towards decisions to breach or follow either norms than items that were more specific to the decision of interest. Therefore, if I were to redo the experiment, I would revise the question about the first statement of the storyboard to "Please rate the appropriateness of the assigned problem", have the student voluntarily justify their work (verbally or in writing, depending on the experimental condition) and add the following question about the second statement of the storyboard: "Please rate the appropriateness of the proposed solution". Along these lines, in the next (and final) section, I propose some directions for future research.

5.4. Potential Directions for Future Research

In my mind, the potential generalization of the theory and method developed in the context of the present study is one of its most exciting take-aways. I write this, of course, not dismissing the importance of the contributions mentioned in section 5.2, but rather excited about the work ahead (for myself and others inspired by this work). One direction for future research would consist of studying how norms influence instructional decisions and how those effects can be measured. In my own thinking about such research, I have recently found inspiration in Cristina Bicchieri's work on social norms. Interestingly, the methods that she employs and describes in her most recent book, *Norms in the Wild* (Bicchieri, 2017), are in some ways similar to the ones I use here. For example, she too uses surveys to measure expectations, although she uses traditional surveys rather than multimedia surveys. She also conducts experiments, but for a different reason than I did: while I conducted an experiment to support the claim that breaching one norm of a given situation might have people abandon their expectations that other norms of that situation will be followed or alter their attitude towards breaches of other norms, she uses experiments to determine whether one's decision to follow (or breach) a norm is caused by their expectations. More specifically, she engages participants in vignettes of situations to which a norm of interest applies that leaves them with the decision of whether to follow or breach that norm. The experimental conditions differ in terms of which types of expectations she primes—expectations that others will follow a norm (which she calls *empirical expectations*[102]) and/or expectations of what others expect them to do (which she calls *normative expectations*). She also administers a number of measures of individual characteristics that she conjectures (and has shown) are predictive of one's decision to breach a norm, including measures of *norm perception*, *norm sensitivity*, *risk perception*, and *risk sensitivity*. She also works to identify the group of people that individuals expected to follow a given norm considers when deciding whether to breach it (which she calls one's *reference network*).

[102] And what I measured with the INR-GCA instrument.

Yet it is worth clarifying that what I am proposing here is not simply the application of general theories of norms to problems in education; instead, I am proposing that we try to apply them to problems in education in order to determine how they will need to be made more subject-specific to suit our purposes (Ball & Forzani, 2007). For example, in the case of the norms of GCA, it is reasonable to expect that the reference group for a geometry teacher includes their students, but this should be empirically confirmed. Second, Bicchieri's measures of risk perception are crafted specifically for each situation under investigation, so a measure of possible perceived risk in breaching, say, the GCA-Figure norm, would need to be developed. Third, I would develop a measure of whether a teacher would follow a given norm of GCA, which could be similar to the *decision instrument* developed by Herbst and colleagues to measure a teacher's tendency to provide the proposition that students will be asked to propose in U.S. high school geometry (vs. allowing students to help choose it; Ko & Herbst, 2020). And last, it would be interesting to know whether teachers and students follow norms of mathematics instruction, such as the GCA-Figure and GCA-Theorem norm, simply because they expect other teachers and students will do the same or because they also believe that people in their reference network expect them to follow those norms and will sanction transgressions.

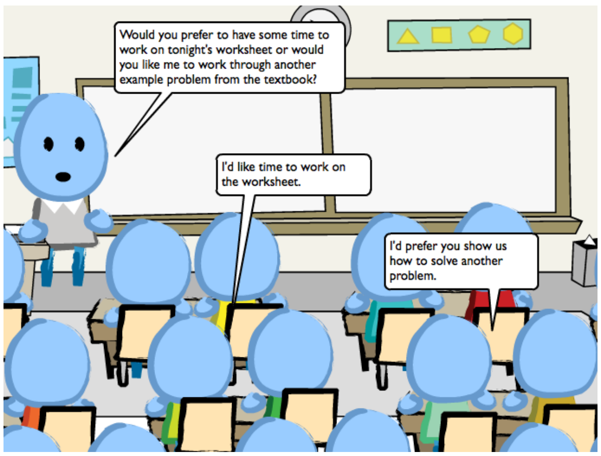
Other interesting and related directions for future research include replicating these experiments with students. The approach to doing so could be similar—having students participate in survey experiments. However, the answers may be very different. For example, it was reasonable to assume that experienced geometry teachers would be aware of its norms, but students typically only take the course once. At what point do they become aware of its norms? Similarly, while the teacher may be part of a student's reference network when deciding to follow or breach a norm, there is reason to believe that their influence on a student's decision may not be the same as the student's influence on their decisions. For example, because teachers are generally seen as an authority in the class (on both mathematical activity and understanding, as well as more general social behaviour), students may perceive larger risks in breaching norms (both subject-specific and general social norms). The question of whether a student breaches a norm in a given scenario because they are not sensitive to risk or because they are unaware of the norm is important because it could contribute to teachers' understanding of students' behaviour similar to how research on student learning has influenced both policy and professional development.

A third fruitful direction for future research may be to consider how different norms that may influence a single decision, including norms that are subject-specific (e.g., situational norms) and general social norms, interact when influencing one's decision. For instance, how do norms of reform instruction suggest that students should be intellectually adventurous square with general social norms of interactions between adults and children, or teachers and students, more specifically?

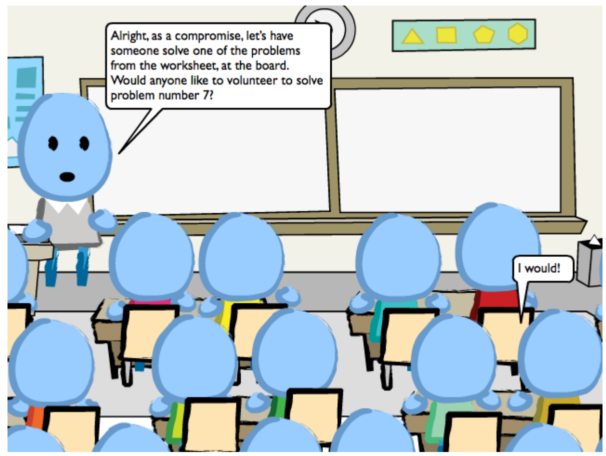
Appendices

APPENDIX A - Sample Storyboard from INR GCA Instrument[103]

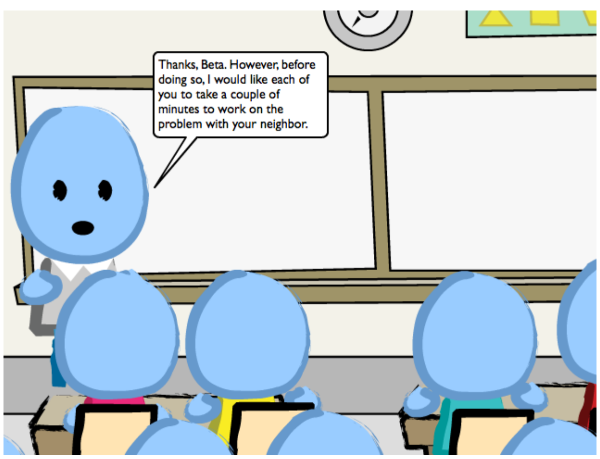
Frame 1



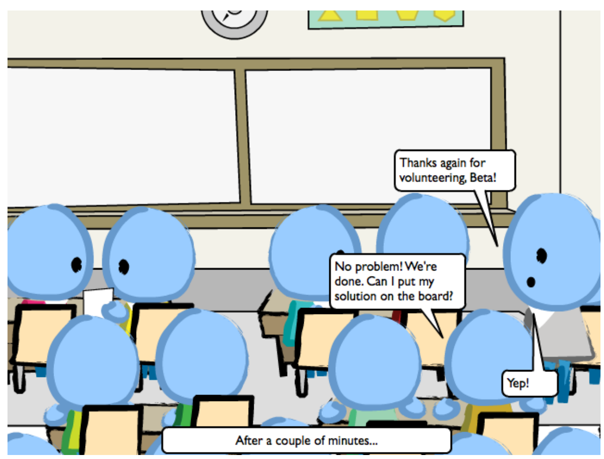
Frame 2



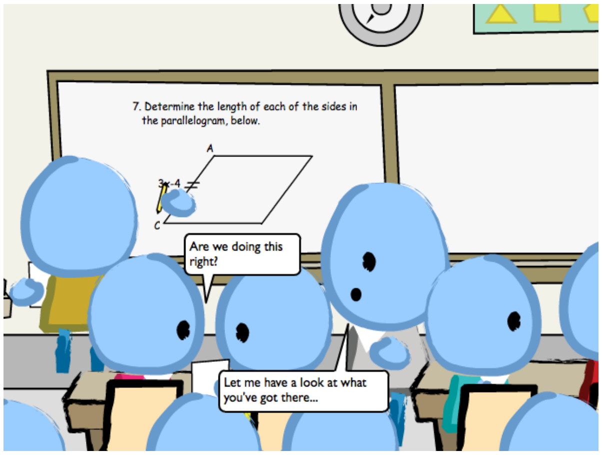
Frame 3



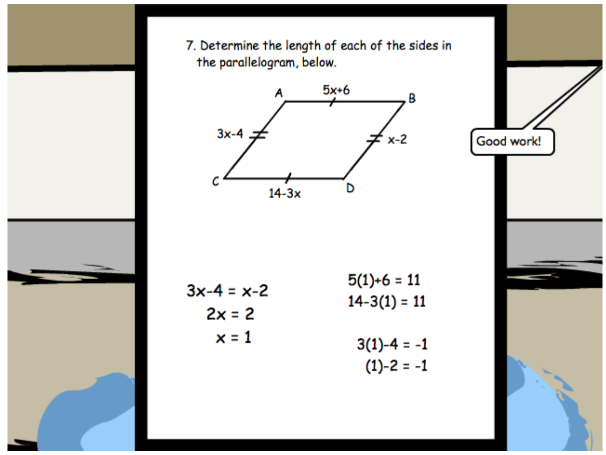
Frame 4



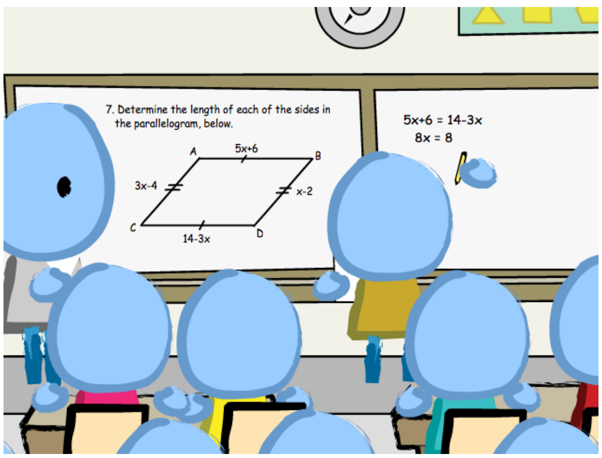
Frame 5



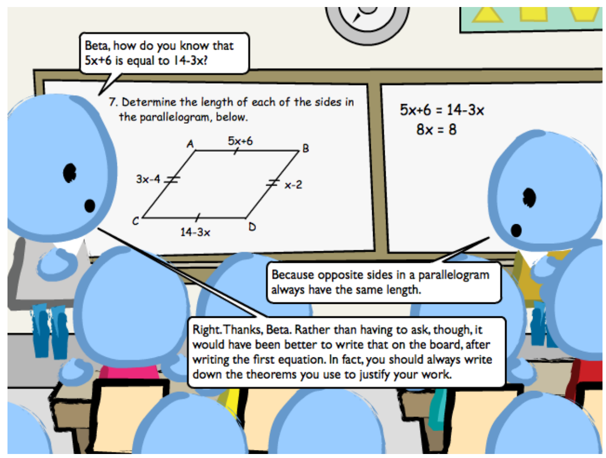
Frame 6



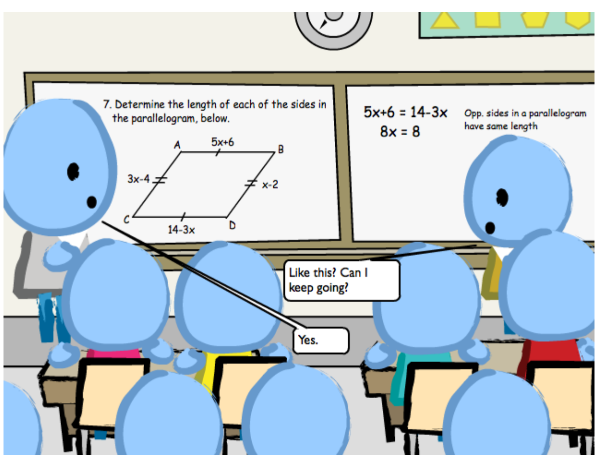
Frame 7



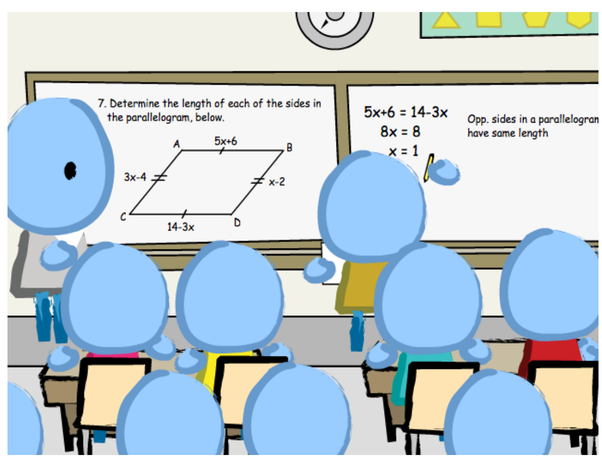
Frame 8



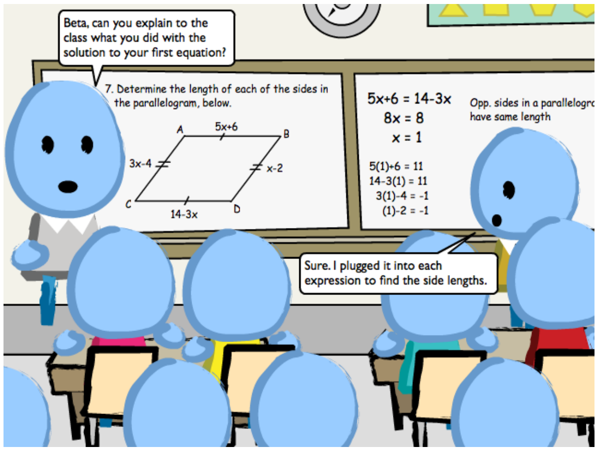
Frame 9



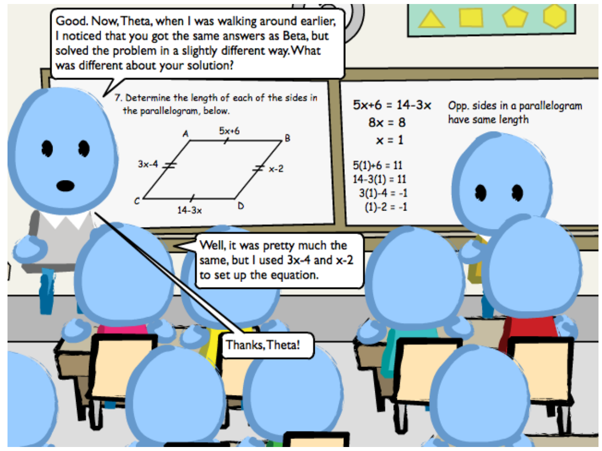
Frame 10



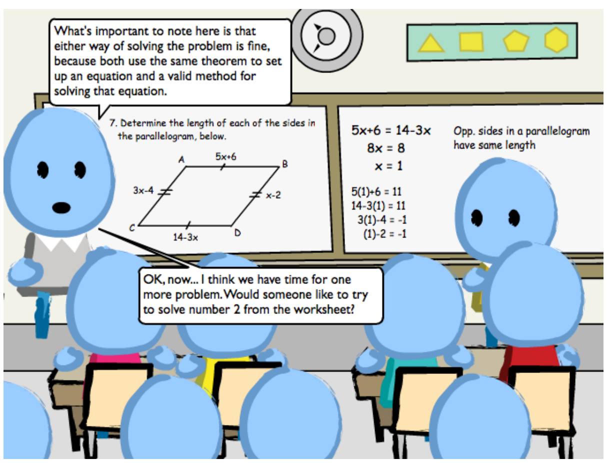
Frame 11



Frame 12

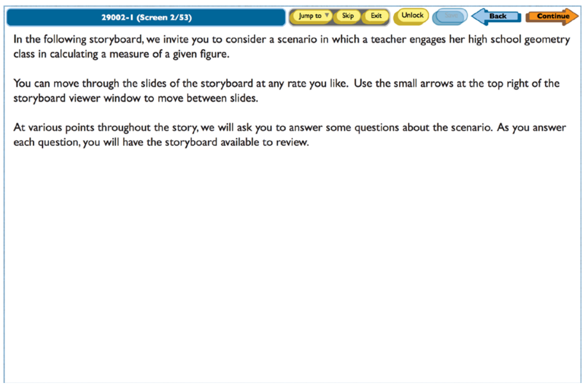


Frame 13

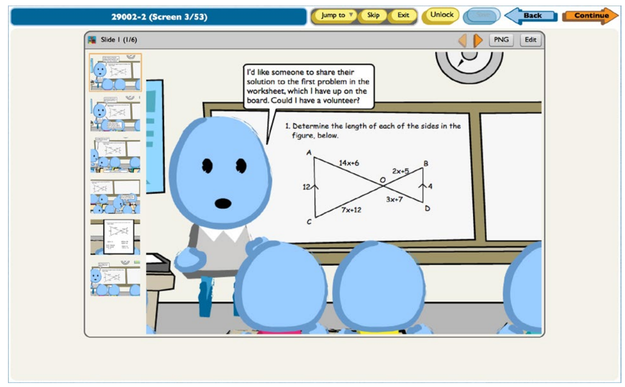


APPENDIX B - Sample Questionnaire from INR GCA Instrument [104]

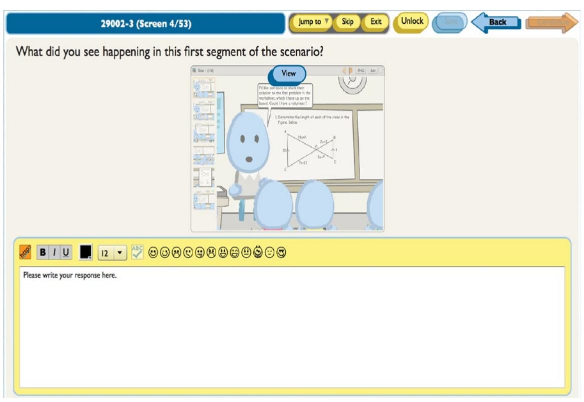
Screen 1



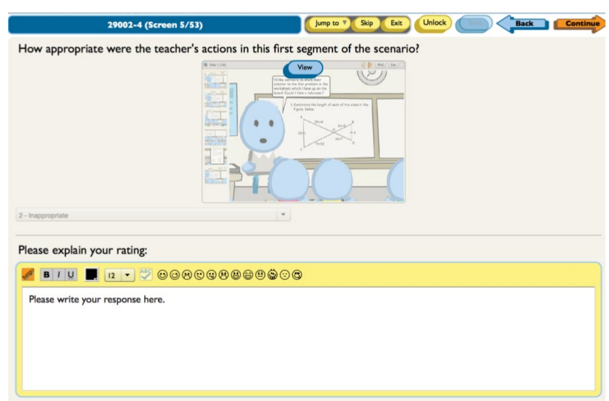
Screen 2



Screen 3

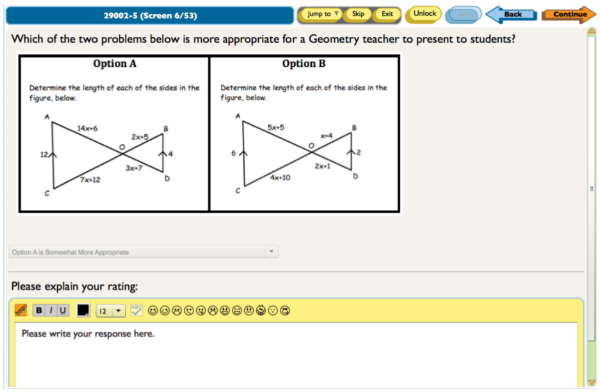


Screen 4

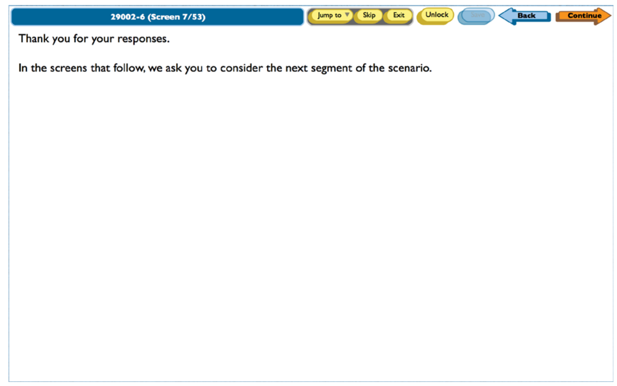


[104] All content and graphics in this Appendix are ©2020, The Regents of the University of Michigan, all rights reserved, used with permission.

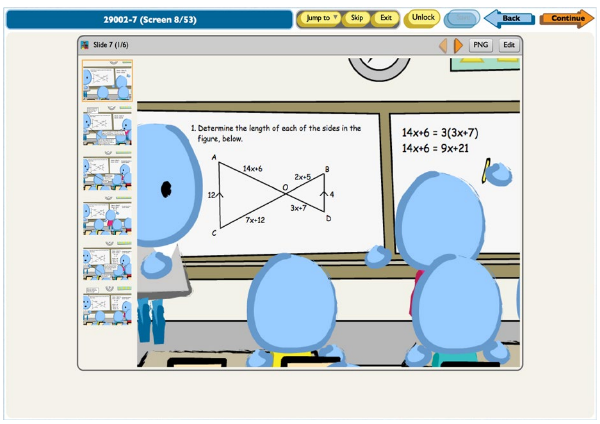
<table>
<tr>
<td>Screen 5



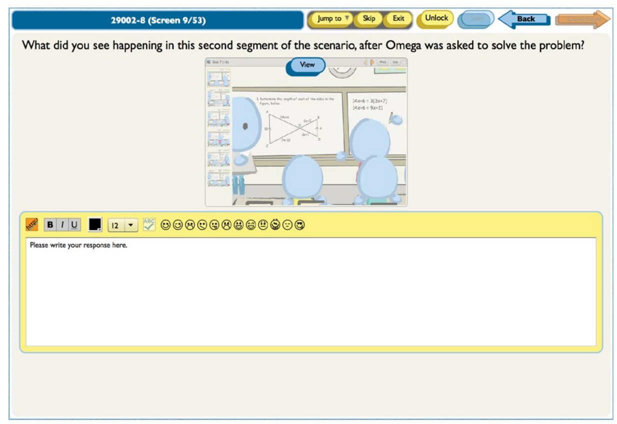
</td>
<td>Screen 6



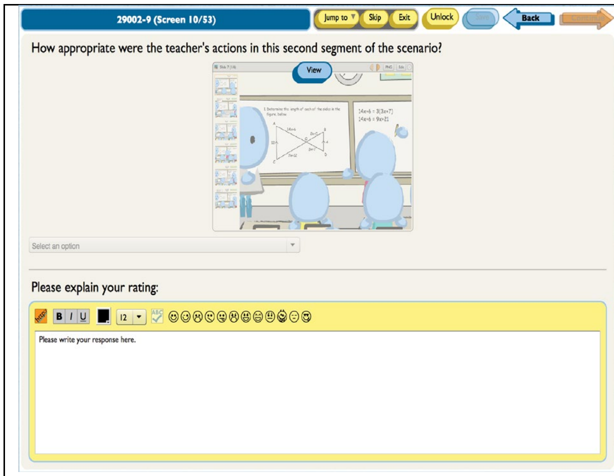
</td>
</tr>
<tr>
<td>Screen 7

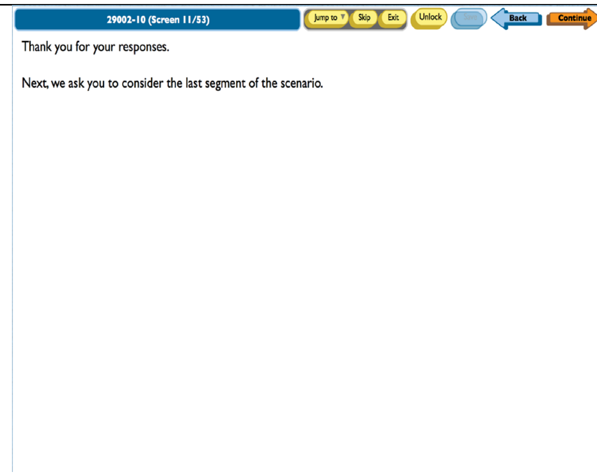


</td>
<td>Screen 8

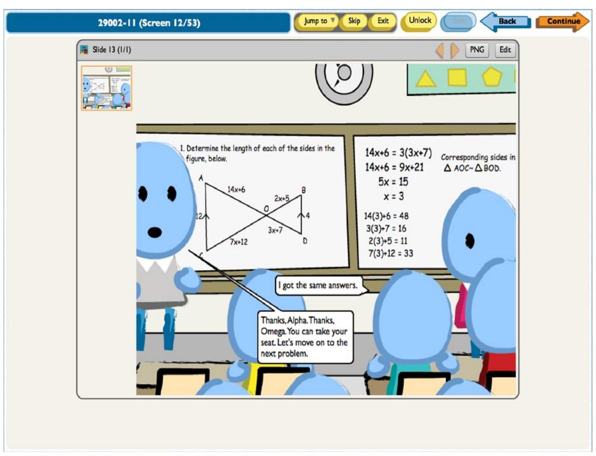


</td>
</tr>
<tr>
<td>Screen 9</td>
<td>Screen 10</td>
</tr>
</table>

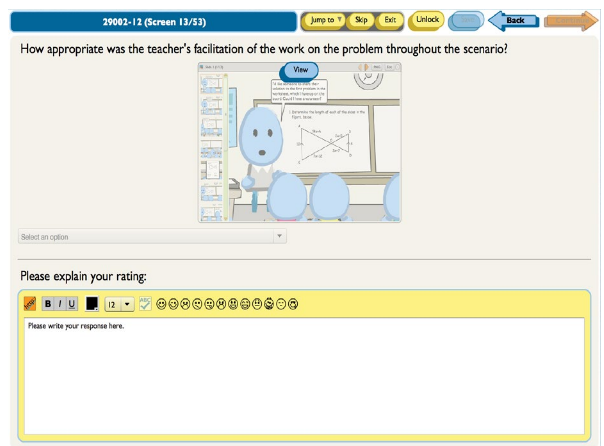




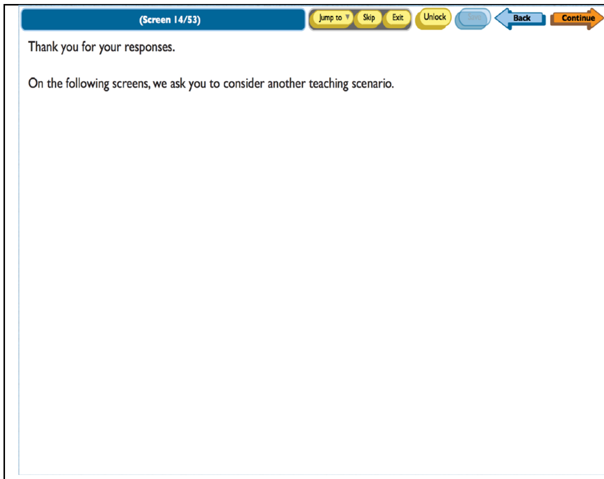
Screen 11



Screen 12



Screen 13


(Screen 14/53)
Jump to
Skip
Exit
Unlock
Back
Continue
Thank you for your responses.
On the following screens, we ask you to consider another teaching scenario.

References

Aleven, V., Koedinger, K. R., Sinclair, H. C., & Snyder, J. (1998, August). Combatting shallow learning in a tutor for geometry problem solving. In *International Conference on Intelligent Tutoring Systems* (pp. 364-373). Berlin, Heidelberg: Springer.

Ambrose, R., Clement, L., Philipp, R., & Chauvot, J. (2004). Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey. *School Science and Mathematics, 104*(2), 56-69.

Association of Mathematics Teacher Educators (AMTE). (2017). *Standards for preparing teachers of mathematics*. Raleigh, NC: Author. Available online at amte.net/standardsBall, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education, 59*(5), 389-407.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *The American Journal of Psychology, 103*(2), 167-193.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education, 59*(5), 389-407.

Behr, M., Lesh, R., Post, T., & Silver E. (1983). Rational Number Concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of Mathematics Concepts and Processes*, (pp. 91-125). New York, NY: Academic Press.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York, NY: Cambridge University Press.

Bicchieri, C. (2017). Norms in the wild: How to diagnose, measure, and change social norms. New York, NY: Oxford University Press.

Boileau, N., Dimmel, J.K., & Herbst, P. G. (2016, November). Teachers' Recognition of the Diagrammatic Register and Its Relationship with Their Mathematical Knowledge for Teaching. In M.B. Wood, E.E. Turner, M. Civil, & J.A. Eli (Eds.), *Proceedings of the Thirty-Eighth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 266-269). Tucson, Arizona.

Boileau, N., & Herbst, P. G. (2015, November). Teachers' Expectations About Geometric Calculations in High School Geometry. In T.G. Bartell, K.N. Bieda, R.T. Putnam, K. Bradfield & H. Dominguez (Eds.), *Proceedings of the Thirty-Seventh Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 269-276). East Lansing, MI.

Boileau, N. & Herbst, P.G. (2016, April). *How Teachers' Evaluate Breaches of Norms in High School Geometry*. Brief research report presented at the Research Pre-session of the Annual Meeting of the National Council of Teachers of Mathematics, San Francisco, CA.

Bourdieu, P. (1990). The logic of practice. Palo Alto, CA: Stanford University Press.

Bourdieu, P. (1998). Practical reason. Palo Alto, CA: Stanford University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brousseau, G. (1981). Problèmes de didactique des décimaux (Recherches en didactique des mathématiques, Vol. 2.1). Grenoble: La pensée sauvage.

Brousseau, G. (1997). Theory of didactical situations in mathematics: Didactique des mathématiques 1970-1990. (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. and Trans.). Dordrecht, The Netherlands: Kluwer.

Brousseau, G., & Warfield, V. M. (1999). The case of Gaël. *The Journal of Mathematical Behavior, 18*(1), 7-52.

Brousseau, G., Brousseau, N., & Warfield, V. (2013). Teaching fractions through situations: A fundamental experiment (Vol. 54). Springer Science & Business Media.

Buchbinder, O., Chazan, D. I., & Capozzoli, M. (2019). Solving Equations: Exploring Instructional Exchanges as Lenses to Understand Teaching and Its Resistance to Reform. *Journal for Research in Mathematics Education*, *50*(1), 51-83

Buchbinder, O., Chazan, D., & Fleming, E. (2015). Insights into the school mathematics tradition from solving linear equations. *For the Learning of Mathematics*, *35*(2), 2-8.

Charmaz, K. (1996). The search for Meaning — Grounded Theory. In J. A. Smith, R. Harré, & L. Van Langenhove (Eds.), *Rethinking Methods in Psychology* (pp.27-49) London, UK: Sage Publications.

Chazan, D., Herbst, P. G., & Clark, L. M. (2016). Research on the teaching of mathematics: A call to theorize the role of society and schooling in mathematics instruction. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1039–1097). Washington, DC: American Educational Research Association.

Chazan, D., Sela, H., & Herbst, P. (2012). Is the role of equations in the doing of word problems in school algebra changing? Initial indications from teacher study groups. *Cognition and Instruction*, *30*(1), 1-38.

Chazan, D., & Sandow, D. (2011). Why did you do that? Reasoning in algebra classrooms. *Mathematics teacher, 104*(6), 460-464.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In M. P. Zanna(Ed.), Advances in experimental social psychology (Vol. 24, pp. 201–234). New York: Academic Press

Cobb, Paul & Gresalfi, Melissa & Hodge, Lynn. (2008). An Interpretive Scheme for Analyzing the Identities That Students Develop in Mathematics Classrooms. *Journal for Research in Mathematics Education, 40*(1), 40-68.

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal, 29*(3), 573–604.

Cobb, P., Yackel, E., & Wood, T. (1989). Young children's emotional acts while doing mathematical problem solving. In D. B. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 117-148). New York: Springer

Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, Instruction, and Research. *Educational Evaluation and Policy Analysis, 25*(2), 119-142.

Cummins, J., Kanold, T., Kenney, M., Malloy, C., & Mojica, Y. (2003). Geometry: Concepts and Applications. New York, NY: Glencoe McGraw-Hill.

Dimmel, J. K. (2015). *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* (Unpublished doctoral dissertation). The University of Michigan, Ann Arbor, MI.

Dimmel, J.K. & Herbst, P. (2015). The semiotic structure of geometry diagrams: How textbook diagrams convey meaning. *Journal for Research in Mathematics Education, 46*(2), 147-195

Dimmel, J.K., & Herbst, P. (2017). Secondary mathematics teachers attitudes' toward alternative communication practices when doing proofs in geometry. *Teaching and Teacher Education, 68*, 151 – 160.

Dimmel, J.K., & Herbst, P. (2018). What details do teachers expect from student proofs? A study of routines for checking proofs. *Journal for Research in Mathematics Education, 49(*3), 261-291.

Dimmel, J. K., & Herbst, P. G. (2020). Proof transcription in high school geometry: a study of what teachers recognize as normative when students present proofs at the board. *Educational Studies in Mathematics*, *105*(1), 71-89.

Dindyal, J. (2007). The need for an inclusive framework for students' thinking in school geometry. *The Mathematics Enthusiast*, *4*(1), 73-83.

Doyle, W. (1977). 4: Paradigms for research on teacher effectiveness. *Review of research in education, 5*(1), 163-198.

Doyle, W. (1983). Academic work. *Review of educational research, 53*(2), 159-199.

Doyle, W. (1988). Work in mathematics classes: The content of students' thinking during instruction. *Educational Psychologist, 23*(2), 167-180

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart and Winston.

Eggins, S. (2004). *Introduction to systemic functional linguistics* (2nd ed.). New York, NY: Continuum.

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching, 15*(1), 13-33.

Furr, R. M., & Bacharach,V. R. (2008). *Psychometrics: An introduction*. Thousand Oaks: Sage Publications.

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction*. Los Angeles: Sage Publications.

Gage, N. L. (Ed.), (1963). *Handbook of research on teaching*. Chicago: Rand McNally.

Garfinkel, H. (1963). A conception of, and experiments with, "trust" as a condition for stable concerted actions. In O. J. Harvey (ed.) *Motivation and Social Interaction*. New York: Ronald Press.

Garfinkel, H. (1964). Studies of the routine grounds of everyday activities. *Social Problems, 11*, 225–250.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Gee, J. P. (2001). Identity as an analytic lens for research in education. In W. G. Se of Research in Education (Vol. 25, pp. 99-125). Washington, DC: American Educational Research Association.

Gibbs, J. P. (1965). Norms: The problem of definition and classification. *American Journal of Sociology*, *70*(5), 586-594.

Glaser, B.G. and Strauss, A.L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.

Goffman, E. (1964). The neglected situation. *American Anthropologist, 66*(6), 133-136.

Gonzalez, R. (2009). *Data analysis for experimental design*. New York, NY: Guilford Press.

González, G. (2013). A geometry teacher's use of a metaphor in relation to a prototypical image to help students remember a set of theorems. *The Journal of Mathematical Behavior*, *32*(3), 397-414.

González, G., & DeJarnette, A. F. (2012). Agency in a geometry review lesson: A linguistic view on teacher and student division of labor. *Linguistics and Education*, *23*(2), 182-199.

González, G. and Herbst, P. (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), 7-33.

González, G., & Herbst, P. G. (2009). Students' conceptions of congruency through the use of dynamic geometry software. *International Journal of Computers for Mathematical Learning*, *14*(2), 153-182.

González, G., & Herbst, P. (2013). An oral proof in a geometry class: How linguistic tools can help map the content of a proof. *Cognition and Instruction*, *31*(3), 271-313.

Green, J. L. (1983). Teaching and learning: A linguistic perspective. *Elementary School Journal, 83*(4), 353-391.

Halliday, M. A. K. (1978). *Language as a social semiotic*. London: Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. London: Edward Arnold.

Harel, G., & Sowder, L. (1998). Students’ proof schemes: Results from exploratory studies. *CBMS Issues in Mathematics Education, 7*, 234-283.

Herbst, P. (2002). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education, 33*(3), 176-203.

Herbst, P. (2003). Using novel tasks to teach mathematics: Three tensions affecting the work of the teacher. *American Educational Research Journal, 4*0, 197-238.

Herbst, P. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education, 37*, 313-347.

Herbst, P. (with González, G., Hsu, H. Y., Chen, C., Weiss, M., and Hamlin, M.). (2010). *Instructional situations and students' opportunities to reason in the high school geometry class*. Manuscript. Deep Blue at the University of Michigan. http://hdl.handle.net/2027.42/78372

Herbst, P., Aaron, W., Dimmel, J., and Erickson, A. (2013, April). *Expanding students' involvement in proof problems: Are geometry teachers willing to depart from the norm?* Paper presented at the 2013 meeting of the American Educational Research Association. Deep Blue at the University of Michigan. http://hdl.handle.net/2027.42/97425

Herbst, P., & Chazan, D. (2003). Exploring the practical rationality of mathematics teaching through conversations about videotaped episodes: The case of engaging students in proving. *For the learning of Mathematics*, 2-14.

Herbst, P., & Chazan, D. (2009). Methodologies for the study of instruction in mathematics classrooms. *Recherches en didactique des mathématiques*, *29*(1), 11-32.

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. (2015). Using multimedia scenarios delivered online to study professional knowledge use in practice. *International Journal of Research and Method in Education*, *38*(3), 272-287.

Herbst, P., & Chazan, D. (2017). The role of theory development in increasing the subject specificity of research on mathematics teaching. In J. Cai (Ed.), *First compendium for research in mathematics education*. Reston, VA: NCTM.

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., Chen, C., Weiss, M., & González, G., with Nachlieli, T., Hamlin, M. and Brach, C. (2009). "Doing Proofs" in Geometry Classrooms. In M. Blanton, D. Stylianou. & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K–16 perspective* (pp. 250–263). New York, NY: Routledge.

Herbst, P., Dimmel, J., and Erickson, A. (2016, April). *High School Mathematics Teachers' Recognition of the Diagrammatic Register in Proof Problems*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington DC.

Herbst, P., Chazan, D., Dimmel, K., Erickson, A., Boileau, N., & Ko, I. (2017, April). *Understanding the rationality of teaching using multimedia questionnaires* [Symposium]. Research Pre-session of the Annual Meeting of the National Council of Teachers of Mathematics, San Antonio, TX.

Herbst, P., & Kosko, K. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. Lo, K. R. Leatham, & L. R. Van Zoest (Eds.), *Research trends in mathematics teacher education* (pp. 23–45). New York, NY: Springer.

Herbst, P., Kosko, K., & Dimmel, J. (2013). How are geometric 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., Shultz., Ko, I., Boileau, N., & Erickson, A. (2018, November). Expanding students' role when doing proofs in high school geometry. In T.E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), *Proceedings of the Fortieth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1028-1035). Greenville, NC.

Hunte, A. A. (2018). Opportunities for Reasoning and Proving in Geometry in Secondary School Textbooks from Trinidad and Tobago. In P. Herbst, U.H. Cheah, P.R. Richard, & K.

Jones (Eds.). International Perspectives on the Teaching and Learning of Geometry in Secondary Schools (pp. 39-58). Springer International Publishing.

Hsu, H.-Y. (2007). Geometric calculations are more than calculations. In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 57-64). Seoul, Korea.

Hsu, H.-Y. (2010a). *The study of Taiwanese students' experiences with geometric calculation with number (GCN) and their performance on GCN and geometric proof (GP)* (Unpublished doctoral dissertation). The University of Michigan, Ann Arbor, MI.

Hsu, H.-Y. (2010b). Conceptualization of the Relationships Between Geometric Proof and Geometric Calculation. In Pinto, M. F. & Kawasaki, T.F. (Eds.). *Proceedings of the 34th Conference of the International Group/or the Psychology of Mathematics Education* (Vol. 3, pp. 105-112). Belo Horizonte, Brazil.

Hsu, H.-Y., & Silver, E. (2014). Cognitive complexity of mathematics instructional tasks in a Taiwanese classroom: An examination of task sources. *Journal for Research in Mathematics Education, 45*(4), 460-496.

Jovanovic, B. D., & Levy, P. S. (1997). A Look at the Rule of Three. *American Statistician*, *51*(2), 137-139.

Kahneman, D. (2011). Thinking, fast and slow. New York, NY: Farrar, Straus and Giroux.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Knuth, E. J. (2000). Student understanding of the Cartesian connection: An exploratory study. *Journal for Research in Mathematics Education, 31*(4), 500-507.

Ko, I. and Herbst, P. (2020). Are teachers amenable to increasing students' scope of work in doing proofs? Estimating teachers' decision-making using a diagnostic classification model. *Journal of Educational Research in Mathematics*, Special Issue, 169-183.

Küchemann, D., & Hoyles, C. (2002). The quality of students' reasons for the steps in a geometric calculation. *Proceedings of the British Society for Research into Learning Mathematics*, *22*(2), 43-48.

Kuntze, S. (2011). Pedagogical content beliefs: Global, content domain-related and situation-specific components. *Educational Studies in Mathematics, 79*(2), 273-292.

Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational evaluation and policy analysis*, *29*(2), 95-114.

Laborde, C. (2005). The hidden role of diagrams in students' construction of meaning in geometry. In *Meaning in mathematics education* (pp. 159-179). New York, NY: Springer.

Lang, B., & Ruane, P. (1981). Geometry in English secondary schools. *Educational Studies in Mathematics*, *12*(1), 123-132.

Lawson, M., & Chinnappan, M. (2000). Knowledge Connectedness in Geometry Problem Solving. Journal for Research in Mathematics Education. 31. 26. 10.2307/749818.

Legros, S., & Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science, 15*(1), 62–80.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, *60*(1), 1-64.

Lesh, R., Post, T. R., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In *Problems of representations in the teaching and learning of mathematics*. Hillsdale, NJ: Erlbaum.

Mann, L. (1970). The Social Psychology of Waiting Lines: The mammoth waiting line is a sophisticated cultural microcosm with a unique set of social rules and behavioral regularities. American scientist, 58(4), 390-398.

Macdonald, H., Zinth, J. D., & Pompelia, S. (2019). 50-State comparison: High school graduation requirements. *Education Commission of The States*.

Martin, D. B. (2000). *Mathematics success and failure among African-American youth*. Hillsdale, NJ: Erlbaum.

Martin, J.R. & White, P.R.R. (2005). *The language of evaluation: Appraisal in English*. London, UK: Palgrave.

Mehan, H., and Wood, H. (1975). The reality of ethnomethodology. Malabar, FL: Krieger.

Milgram, S., Liberty, H. J., Toledo, R., & Wackenhut, J. (1986). Response to intrusion into waiting lines. Journal of Personality and Social Psychology, 51(4), 683–689.

Moschkovich, J., Schoenfeld, A., & Arcavi, A. H. (1993). Aspects of understanding: On multiple perspectives and representations of linear relations and connections among them. In T. A. Romberg, E. Fennema, & T. P. Carpenter (Eds.), *Integrating research on the graphical representation of functions* (pp. 69-100). Hillsdale, NJ: Erlbaum.

Much, N. C., & Shweder, R. A. (1978). Speaking of rules: The analysis of culture in breach. *New Directions for Child and Adolescent Development, 1978*(2), 19-39.

Nachlieli, T. (2011). Co-facilitation of study groups around animated scenes: The discourse of a moderator and a researcher. *ZDM*, *43*(1), 53-64.

Nachlieli, T., Herbst, P., & 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*, 427-459.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to Actions: Ensuring Mathematical Success for All*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2018). *Catalyzing Change in High School Mathematics: Initiating Critical Conversations*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: [National Governors Association Center for Best Practices (NGA Center)]/[Council of Chief State School Officers (CCSO)].

Otten, S., Males, L. M., & Gilbertson, N. J. (2014). The introduction of proof in secondary geometry textbooks. *International Journal of Educational Research*, *64*, 107-118.

Rawls, A. W. (2011). Harold Garfinkel. In Ritzer, George (Eds.), *The Blackwell Companion to Major Social Theorists*. London, UK: Blackwell.

Reid, D. A., & Knipping, C. (2010). *Proof in mathematics education: Research, learning, and teaching*. Rotterdam, The Netherlands: Sense.

Schleppegrell, M. J. (2012). Systemic Functional Linguistics. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 21-34). London: Routledge.

Schoenfeld, A. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. Grows (Ed.), *Handbook for Research on Mathematics Teaching and Learning* (p.334-370). New York, NY: MacMillan.

Schoenfeld, A. H. (2010). *How We Think: A Theory of Goal-Oriented Decision Making and its Educational Applications*. New York, NY: Routledge.

Schumann, H., & Green, D. (2000). New protocols for solving geometric calculation problems incorporating dynamic geometry and computer algebra software. *International Journal of Mathematical Education in Science and Technology, 31*(3), 319-339.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company.

Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York, NY: Routledge.

Shulman, L.S. (1986). Paradigms and research programs for the study of teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Maemillan.

Shultz, M., Herbst, P., Boileau, N., & Erickson, A. (2018, April). *Measuring geometry teachers' willingness to expand proof practices* [Paper Session]. Annual Meeting of the American Educational Research Association, New York, NY.

Stacey, K., & MacGregor, M. (1999). Implications for mathematics education policy of research on algebra learning. *Australian Journal of Education, 43*(1), 58-71.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488.

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. Educational Psychologist, 35(2), 87-100.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan, 79*(1), 14-21.

Stigler, J. W., & Hiebert, J. (1998). Teaching Is a Cultural Activity. *American Educator, 22* (4), 4-11.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*(2), 213-226.

Strauss, A., & Corbin, J. 1994. Grounded theory method. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp.273-285). Thousand Oaks, CA: Sage.

Stylianou, D. A. (2011). An examination of middle school students' representation practices in mathematical problem solving through the lens of expert work: Towards an organizing scheme. *Educational Studies in Mathematics, 76*(3), 265-280.

Suwito, A., Yuwono, I., Parta, I. N., & Irawati, S. (September, 2016). The Linkage of Problem Solving Between Geometry and Algebra: What Is Their Correlation? In S. Saptono, M.

Masturi, A. Purwinarko, & D. Alighiri (Eds.), *Proceedings of the international conference on mathematics, science, and education* (pp. M-22-M-27). Semarang, Indonesia.

Tripathi, P. N. (2008). Developing Mathematical Understanding through Multiple Representations. *Mathematics Teaching in the Middle School, 13*(8), 438–445.

Voigt, J. (1995). Thematic patterns of interaction and sociomathematical norms. In P. Cobb & H. Bauersfeld (Eds.), *Emergence of mathematical meaning: Interaction in classroom cultures* (p. 163–201). Hillsdale, NJ: Erlbaum.

Weiss, M., Herbst, P., & Chen, C. (2009). Teachers' perspectives on "authentic mathematics" and the two-column proof form. *Educational Studies in Mathematics*, *70*(3), 275-293.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal of Research in Mathematics Education, 22*, 390-408