

Investigating Students' Proof Reasoning: Analyzing Students' Oral Proof Explanations and their Written Proofs in High School Geometry

Michael L. Winer^{1*} , Michael T. Battista² 

¹Eastern Washington University, USA

²The Ohio State University, USA

*Corresponding Author: mwiner@ewu.edu

Citation: Winer, M. L., & Battista, M. T. (2022). Investigating Students' Proof Reasoning: Analyzing Students' Oral Proof Explanations and their Written Proofs in High School Geometry. *International Electronic Journal of Mathematics Education*, 17(2), em0677. <https://doi.org/10.29333/iejme/11713>

ARTICLE INFO

Received: 21 Sep. 2021

Accepted: 28 Jan. 2022

ABSTRACT

Constructing formal geometry proofs in is an important topic in the mathematics curriculum. But students' difficulties with proof are well documented. This article focuses on proofs that use triangle congruence postulates in US high school geometry. Examining students' proof reasoning in one-on-one task-based interviews, we analyzed students' oral planning/explanations and two-column written proofs. We found that much of the reasoning students conveyed in their oral planning/explanations was not incorporated into their written proofs, resulting in formalization and fatal logical reasoning gaps. We found that the majority of the students utilized sound reasoning in their oral explanations, but struggled to capture that reasoning in their written two-column proofs. We link our findings to learning progression research, further elaborating the van Hiele levels. Our research contributes to understanding the learning and teaching of proof and to the ongoing controversy over the use of two-column proofs in the high school geometry curriculum.

Keywords: geometry proofs, two-column proofs, learning progressions, student cognition, mathematics education

INTRODUCTION

Establishing the validity of reasoning is critical in understanding, creating, and using mathematics, and the standard way of establishing mathematical validity is proof. Thus, researchers have argued that proving is essential to students' mathematics learning, and it is considered an important topic in the mathematics curriculum (Battista & Clements, 1995; Cirillo & Herbst, 2012; Common Core State Standards Initiative, 2010; Hanna & Jahnke, 1996; Herbst, 2002; National Council of Teachers of Mathematics, 1989, 2000; Stylianides, 2007; Wu, 1996). In the US, most instruction on proof is limited to high school geometry. However, research conducted on students' proof performance has found that the majority of high school geometry students struggle with this topic (Chazan, 1993; McCrone & Martin, 2004; Senk, 1985, 1989). For instance, Senk (1985) found that only 30% of US students in a full-year geometry course that covered proof reached a 75% mastery level of proof writing. McCrone et al. (2002) found that "[S]tudents performed poorly on items that required them to write a formal proof with no support. They also had difficulty on items that required students to make a single deduction from a given piece of information" (p. 1).

In the US, the most common context for formal geometry proof is the two-column format (Herbst, 2002; Pair et al., 2021), which was created to make proof accessible to high school students (Herbst, 2002). Even though the two-column proof format continues to be widely used in US high school geometry textbooks and classrooms (Kelley, 2013; Stylianides et al., 2017), the use of such proofs is controversial. In a survey of recent mathematics education conference attendees, Pair et al. (2021) found that 79% of respondents stated that there is value in two-column proofs and 43.5% said that they were probably or definitely not in favor of eliminating two-column proofs. In contrast, 21% of respondents did not see the value of such proofs, and 35.5% probably or definitely thought that such proofs should be eliminated. Further, a number of mathematics education scholars have criticized the format. For instance, Schoenfeld (1988) criticized two-column proof as focusing on form over mathematical substance. As an example, he provides a proof of the Base Angles Theorem for Isosceles Triangles that he claims most mathematicians would find adequate, but that would not be accepted in many high school geometry classrooms because it omits details required in the two-column format. More generally, according to Hanna (1989), because mathematicians present mathematical results in the form of theorems and proofs, this practice is mistakenly seen as the core of mathematical justification (Lakatos, 1976). However, mathematicians generally discover new ideas using intuitive and empirical methods. In creating mathematics, they pose

problems, analyze examples, make and revise conjectures, search for counterexamples—it is only after this intuitive “playing around” has intuitively convinced them that something might be true that sophisticated users of mathematics turn to verifying their conclusions using a step-by-step deductive proof. Furthermore, Cirillo et al. (2021), with Herbst and Miyakawa (2008), argue that most of what students do in high school geometry proofs is framed by Givens-Prove-Figure Template/Configuration (GPFT) instead of the more genuine practice of proving statements in theorem format, thus masquerading the nature and role of mathematical proof.

In contrast to naysayers, Weiss et al. (2009) claim that the two-column format can productively engage students in proof by offering constraints and supports for constructing valid proofs. The two-column format can aid both teachers and students in reflecting on and evaluating the validity of deductive proofs (Herbst, 2002). Additionally, there are three counterarguments to the claim that two-column proofs promote form over substance. First, form, with its sequential structuring, is a critical part of a deductive argument. So, it is more likely that superficial instructional focus on form is the problem, not form itself. Second, a confounding issue is the common instructional practice of posing geometry problems in which students can assume that the conclusion is true (Herbst, 2002; Mariotti, 2006), a practice that emphasizes proof formality over proof as convincing justification. This issue can to a great extent be alleviated in more exploratory approaches to geometry instruction such as those utilizing DG in which students have to decide whether to prove an empirical finding or find a counterexample to its generality (Battista, 2007; Gilbertson et al., 2013; Mariotti, 2006). The third confounding issue is posing all proof problems in GPFT two-column format, a practice that can be alleviated by having students prove *theorems* in both two-column and paragraph forms, and explicitly attending to what makes a proof valid.

Given the importance of formal proof in mathematics and the widespread use of the two-column format, it is important for research to go beyond paper-and-pencil evaluations of its correct use to investigate students’ reasoning as they engage in the proof process. Numerous researchers have stated that more in-depth research is needed on students’ development of geometric proof understanding, reasoning, and skills (Battista, 2007; Cirillo & May, 2020; Sinclair et al., 2017). Thus, the current investigation conducted individual interviews of students as they orally planned proofs then wrote them in two-column format while thinking aloud. By including students’ oral reasoning along with their written proofs, the present study provides an additional perspective on whether and how two-column proofs can be useful in developing students’ proof reasoning, and it identifies elements of this format that cause student difficulties.

Research on Oral Versus Written Proofs

In investigating students’ development of proof reasoning, a number of researchers have contrasted oral versus written proofs. Focusing on 8th grade number theory proofs, Campbell et al. (2021) found that students often struggled to write proofs that they presented orally, and “found no evidence that students’ written arguments were more sophisticated than their oral arguments” (p. 21). They claimed that some research suggests that students produce more sophisticated written proofs, but other research indicates that students’ oral proofs are of higher quality. Soto-Johnson and Fuller (2012) studied oral and written proofs of college students in abstract algebra, finding that some students who understood the ideas needed for a proof “could not write a coherent proof.... [but] these same students provided fairly sound verbal proofs” (p. 1). In another study focusing on pattern problems for 14-15-year olds, Stylianides (2019) found that oral rather than written proofs of patterns were more likely to meet proof standards. In this case, the written proofs were produced in small group work and afterwards oral arguments were presented in front of the class.

However, even though past *oral/written* studies provide valuable insight into the development of students’ proof reasoning, none of them has dealt with the two-column proof format in geometry. Given the widespread use of the two-column format, it is important to investigate both *oral and written* components of proof in this context. Thus, the present study, focusing on proofs that use the triangle congruence postulates, investigated students’ proof reasoning in (a) orally stated proof plans, (b) subsequent written two-column proofs, and (c) think-alouds during proof writing. The research questions are explained in the following.

Research Questions

According to Mariotti (2006), two of the three main streams of research about proof, both of which we address, are: “What is the status of proof at school?” and “What are the main difficulties that students face in relation to proof?” (p. 174). And there are two main approaches to investigating students’ conceptions of proof, again, both of which we address: (1) taking a broad perspective, researchers analyze the different ways of thinking that can be described and classified as proof; and (2) starting from an epistemological position that bestows a specific status to mathematical formal proofs, researchers describe and explain difficulties and obstacles encountered by students in their attempts to construct formal proofs. More specifically, our research questions are as follows:

1. How do students, in US geometry classes in which a two-column format is prevalent, understand the structure of proof, orally when planning, and during the process of writing proofs in two-column format?
2. In which form of expression, oral or written, are students less likely to make errors and produce valid proofs/defensible proof reasoning?
3. What kinds of errors do students in the two expression forms, oral and written, make?
4. What aspects of two-column proofs do students struggle with?

CONCEPTUAL FRAMEWORK

Taking a psychological constructivist perspective, we believe that students must personally construct mathematical ideas as they repeatedly cycle through phases of action, reflection, and abstraction, and that effective mathematics instruction carefully guides and supports students' construction of personally meaningful mathematical concepts and ways of reasoning based on knowledge of student reasoning. Indeed,

There is a good deal of evidence that learning is enhanced when teachers pay attention to the knowledge and beliefs that learners bring to a learning task, use this knowledge as a starting point for new instruction, and monitor students' changing conceptions as instruction proceeds (Bransford et al., 1999, p. 11).

In this article, we carefully examine students' oral and written proof reasoning in a way that can be used to reflect on how best to support the development of students' proof understanding and reasoning.

In addition, we elaborate our research in a way that is instructionally useful by integrating it with learning progressions in geometry. According to the National Research Council (2007), "Learning progressions (LP) are descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic" (p. 214). An LP for a topic (a) starts with the informal, pre-instructional reasoning typically possessed by students; (b) ends with the formal mathematical concepts targeted by instruction; and (c) indicates cognitive plateaus reached by students in moving from (a) to (b) (Battista, 2011, 2012). Learning progressions are playing an increasingly important role in mathematics and science education (National Research Council, 2001, 2007; Smith et al., 2006). They are strongly suggested for use in assessment, standards, curriculum design, teaching, and research on learning and teaching (Sztajn et al., 2012).

Types of Cognitive Structuring Used to Reason About Geometric Proofs

Axiomatic proof is a mathematical argument consisting of a sequence of connected statements in support of a mathematical claim, each statement logically deduced from previous statements and justified by a combination of given statements, axioms, and previously proven theorems (Beman & Smith, 1899; Herbst, 2002; Movshovitz-Hadar, 2001). In the context of an axiomatic system, deduction is comprised of three main components: a set of premises, a conclusion, and a justification (Anderson et al., 1985; Duval, 2007). The *premises* are a set of propositions or conditions that are to be assumed to be true or correct. A *conclusion* is a proposition stating the outcome of a logical deduction. A *justification* is a sequence of deductions and citations of axioms and previously proved theorems that is required to draw the conclusion. A (correct) valid *deduction* or valid inference¹ occurs when a person uses a sufficient sequence of assumed/proven premises and the rules of logic to draw a logically certain conclusion that is justified within a specified axiomatic system (Cirillio & Hummer, 2021; Movshovitz-Hadar, 2001).

The reasoning involved in constructing geometric proofs is complex and involves three types of cognitive structuring: spatial, geometric, and logical/axiomatic (Battista, 2008). *Spatial structuring* mentally constructs a spatial organization or form for a physical/pictorial/graphic object or set of objects (Battista, 1999, 2008). *Geometric structuring* consists of using formal geometric concepts and properties to describe the interrelationships between components that determine a geometric object's spatial structure. Examples of formal concepts that are included in geometric structuring are congruence, parallelism, slope, length, angle measure etc. For a geometric structuring of an object to make sense to a person, it must evoke an interiorized appropriate spatial structuring of the shape as well as interiorized formal geometric concepts used to describe the shape (Battista, 2008).

To begin creating a geometric proof of a statement about geometric relationships—with a given, or student drawn diagram—students must first construct linked spatial and geometric structuring. However, to create a formal geometric proof, not only does the student need an appropriately linked spatial and geometric structuring, but the student must link this spatial and geometric structuring to the third and fourth types of structuring—logical and axiomatic structuring.

Logical structuring in a proof is the process of making a series of deductions assumed to be consistent with the rules of logic in an attempt to prove the desired conclusion from the given premises. Correct logical structuring occurs when a student's deductions are not only consistent with the rules of logic, but organized in an appropriate sequence to prove the desired conclusion from the given premises. In the two-column geometry proof format, logical structuring is the sequence of conclusions/statements that one deduces and lists in the left-hand column. One major error that can occur in a proof's logical structure is when the argument has a gap in it (Duval, 2007). A *reasoning gap* occurs when a student deduces a conclusion by applying an axiom or theorem whose premises have not been explicitly established by the given conditions or previous deductions in their proof.

Axiomatic structuring in a proof is the process of explicitly situating and justifying deductions and logical structuring within a given axiomatic system. In the two-column proof format, one can think of axiomatic structuring as the sequence deductions in the left-hand column linked to the required justifications in the right-hand column. A two-column proof exhibits a correct axiomatic structure if every one of its conclusions exhibits a logically valid deduction (correct logical structuring) and each conclusion is correctly justified in the right-hand column by an appropriate axiom or theorem from the axiomatic system. So, the only feature that distinguishes logical and axiomatic structuring is that the latter requires correct, explicit axiomatic justifications. Of course, both logical and axiomatic structuring can occur even if a proof is not in two-column format (although they may be a bit more difficult to discern).

¹ Logical deduction, deduction, logical inference, and deductive step will all be used interchangeably throughout the article.

Using Logical Deductions to Construct Geometric Proofs

A major component of the validity of students' geometric proofs is the sequence of logical deductions he or she makes. Duval (2007) argued that the most important part of constructing a logical proof is understanding the status or function of each proposition within a single deduction and within a chain of deductions. Duval (2007, p. 140) defined status as "the specific function, the particular role of each proposition within the set of the other propositions which are required or stated to get a proof or to produce an argumentation." Each proposition within a single deduction serves one of the following three functions: a premise, a conclusion, or a justification (Clements & Battista, 1992; Duval, 2007; Wertheimer, 1990). Deductions within a proof are *logically linked* when the student takes conclusions from earlier deductions and uses them as premises for future deductions (Duval, 2007; Mariotti, 2006; McCrone & Martin, 2004). In other words, once a student has deduced a conclusion he or she needs to change its status to a premise so he or she can use it in deducing a future conclusion for their proof. A *complete logical chain of deductions for a proof* is a sequence of conclusions that consists of logically linked deductions that originate from the givens and end with the desired conclusion.

METHODS

Sample and Research Design

To capture students' mathematical reasoning, Winer conducted a series of one-on-one semi-structured task-based interviews (Goldin, 2000) with seven ninth and tenth grade geometry students who were asked to complete a series of proof problems. All participants were volunteers who were currently enrolled in a proof-based geometry course in which they had already completed a unit on triangle congruence proofs. The participants were from a suburban high school located in the Midwest United States. The participants came from two different geometry classes taught by two different teachers: five participants (four girls and one boy) came from an Honor's geometry class while the remaining two participants (two girls) came from a normal-track geometry class. Students were each individually interviewed for five one-hour sessions in which they worked on 12 proof problems. For each problem, students were first asked to *orally plan*² their proof and only after they had described their proof plan orally were they asked to write out their formal proof. In addition, as students wrote their proofs they were asked to explain their reasoning aloud to provide additional insight into their reasoning. Students were given a reference sheet with theorems and axioms on it, which acted as the default axiomatic system. The reference sheet was created using the theorems and axioms that were found in the student's textbook so as to match the way that theorems and axioms were stated and named in their classes. If a student wanted to use a theorem that was not present on the reference sheet they were asked to prove why that theorem was valid, given what was included in the reference sheet. To cover the types of proofs that were typically found in high school geometry, all proof problems were selected or adapted from mainstream high school geometry textbooks (e.g., Larson et al., 2008) or relevant research studies on geometric proofs (e.g., Anderson et al., 1981; Chen & Herbst, 2013; Heinze, 2008; Martin & McCrone, 2003; Senk, 1985). Problems were posed in GPFT format, in theorem format with no diagram, and in diagram format in which auxiliary lines were needed. Including this format variety permitted a more general examination of students' overall geometry proof reasoning.

Data Sources and Data Analysis

This study is a generative qualitative research study in which the theory is created from empirical data and previous research. Clement (2000) stated, "The purpose of a generative study is to generate new observation categories and new elements of a theoretical model in the form of descriptions of mental structures or processes that can explain the data" (p. 558). During implementation, the interviewer collected data from multiple sources, such as two video recordings for each interview session (from different angles), students' written work, and audio journals created by the interviewer after each interview session, enabling triangulation during data analysis. All video recorded interviews were transcribed and later analyzed with the other data sources using the constant comparative method from Grounded Theory (Glaser & Strauss, 1967) and retrospective analysis (Steffe & Thompson, 2000).

After transcriptions were completed, they were segmented into smaller episodes that correspond to each proof problem. The segmenting of episodes allowed us to know when the student began and ended a proof problem as well as to separate the interview sessions into more feasible units of analysis. The transcriptions not only represented students' oral utterances, but also presented any relevant actions (gestures, written responses, body language, stated and described when students made demarcations on diagrams, etc.) that occurred during the interviews.

For each problem episode, we began the data analysis in an open coding procedure from the constant comparative analysis method (Glaser & Strauss, 1967), in which we identified and annotated interesting or significant incidents of student reasoning (Glaser, 1978; Glaser & Strauss, 1967). This analysis was an iterative one in which incidents and subsequent analytical commentaries were compared with similar incidents found in other episodes across participants and were constantly revised and modified. Once the analytical commentary and codes for each episode stabilized and were tentatively viable, we returned to data and did a more rigorous search for common trends and patterns among the students' reasoning with particular attention to both what the students orally stated as they planned and constructed their proof to what they wrote in the formal proof. When searching for trends, we again used a constant comparative method to identify, construct, and refine conceptual categories that emerged from the data. The conceptual categories were also developed using an iterative cycle of constructing, criticizing, and

² Students were allowed to write down notes and mark the diagram. But they were asked to not write their formal proof until they had completed their oral plan unless otherwise noted.

revising. This iterative cycle and constant comparison analysis continued until the categories became theoretically saturated (Glaser & Strauss, 1967). The conceptual categories identified and developed from this process were two types of reasoning gaps in the students' proof reasoning, which are defined and discussed in the findings section.

FINDINGS

When we refer to a student's *written/formal proof*, we focus only on what the student wrote in their two-column proofs. When we refer to a student's *overall proof reasoning*, we consider and analyze three components of data: students' oral proof plan, students' oral explanations during their proof writing, and their written proofs. Our analysis of these three data components indicated that students committed two types of reasoning gaps in their overall proof reasoning, formalization gaps and fatal logical gaps.

In the following sections, we define these two types of reasoning gaps and provide examples of student work that falls into each one of the gap categories. We also provide an analysis of the possible causes for these gaps. As we illustrate the examples of the two types of gaps, due to length restrictions, we will not address every gap that is found in each of the examples of the students' proofs, but only ones that illustrate how we coded the gaps. We follow the descriptions and student example gaps with a quantitative analysis of the frequency of these gaps as they occurred throughout the interviews. All students' names are pseudonyms.

Formalizations Gaps

A *formalization gap* (FG) in a student's proof reasoning occurs when the student is missing critical details and/or deductions in their written proof, but made oral statements and/or gestures in their plans or while working on their written proof that indicated some intuitive or informal understanding related to the missing information. We coded an instance as a formalization gap if it fit one of the following three categories.

Formalization gap 1 (FG1)–Curtailement

Formalization gap 1 (FG1)–Curtailement occurred when students stated correct deductions orally in their proof planning or while they wrote their formal proofs, but did not write all of these deductions in their formal proof. This category included instances in which the students orally explained and justified a correct series of deductive steps, but combined these steps into a smaller set of written deductions in their formal proofs causing their formal proofs to be missing steps.

A formalization gap was coded as one instance when students were missing steps between two deductions in their written proof. The length of a reasoning gap refers to the number of missing deductive steps that are needed to logically link the two written deductions. For example, if a reasoning gap between two deductions has a length of four, it means that there are four missing deductive steps needed to logically close the gap between the deductions. The FG1 gap is illustrated with Rose's oral explanations and formal proof for Problem M (see [Figure 1](#)).

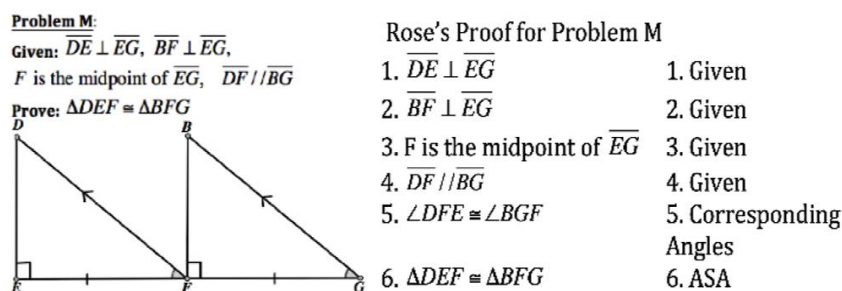


Figure 1. Rose's demarcations of the given diagram and written proof for Problem M

During oral planning:

Rose: So, DE is perpendicular to EG is down here [marks Angle DEF as a right angle on the given diagram]. And then the same thing with BF [marks Angle BFG as a right angle]. ...

IN: And how did you know they were right angles?

Rose: I was marking for perpendicular. *If it is perpendicular it has to be a right angle*³. So, F is the midpoint of EG, which means that since it is the midpoint these [points to Segment EF and FG] have to be the same length [marks Segment EF and FG as congruent]. And then DF is parallel to BG. So, [marks Segments DF and BG as parallel] I drew parallel signs. And DEF and BFG [outlines triangles DEF and BFG with her finger], so proving that those two triangles [are congruent]. [Pause]. So, since this [points at Segment EG] is a straight line and these [outlines Segments DF and BG] are parallel they intersect *at the same angle* [makes X gesture with her hands] which makes these [marks Angles DFE and BGF as congruent]

³ We use italics to emphasize the points we are commenting about in the analysis.

Corresponding Angles. And I just proved that the triangles because Angle-Side-Angle [points to Angle BGF, Segment FG, and Angle BFG].

Rose's reasoning exhibited an FG1 Curtailment gap of length three between steps 5 and 6. First, in Rose's oral plan/explanations, she clearly demarcated that Angles BFG and DEF were right angles on the diagram. When the interviewer asked how she knew that those angles were right angles, she replied "If it is perpendicular it has to be a right angle." From this oral statement and her demarcation of the right angles on the diagram we have evidence that Rose understands that the perpendicular given condition implies that Angles BFG and DEF are right angles. Later, however, Rose did not write this oral deduction involving right angles and perpendicular segments in her formal proof. After Rose had finished writing her formal proof, in response to a query about the right angles from the interviewer, she replied that it was "self-explanatory, given that they are perpendicular they would have to be right angles." So, Rose appeared to not write that these angles were right angles because she seemed to assume that the conclusion was intuitively obvious. Dreyfus and Hadus (1987), commenting on proof principles, stated that "Even 'obvious' statements must be proved" (p. 48).

Second, Rose's formal proof is also missing the deduction that Angles BFG and DEF are congruent by the All Right Angles are Congruent Theorem. However, Rose evidenced in her oral planning that she understood the need for this conclusion when she explicitly gestured to Angle BGF, Segment FG, and Angle BFG (i.e., a right angle) on the diagram to represent the congruent parts in justifying by ASA Triangle Congruence that Triangles BFG and DEF were congruent. From her gesture to Angle BFG, which she had demarcated as a right angle on the diagram (see Figure 1), it can be inferred that Rose used the notion that the two right angles were congruent to satisfy the premise of second pair of congruent angles needed to apply the ASA Postulate. This shows that Rose had valid informal reasoning in her oral planning of the proof, but she did not formalize it correctly in her written proof—she simply skipped steps. There were many other instances of students not writing a deduction that concluded that right angles are congruent for problems in which the students used them as a premise to implement a later deduction, typically involving triangle congruence. However, in all of these instances the right angles were demarcated on the diagram by the student.

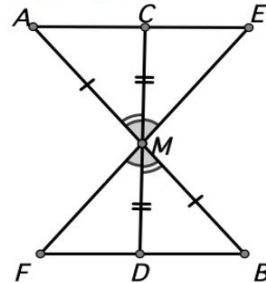
Third, is Rose's missing deduction in her written proof that Segment EF is congruent to Segment FG by the definition of midpoint. Rose stated that "So, F is the midpoint of EG, which means that since it is the midpoint these [points to Segment EF and FG] have to be the same length [marks Segment EF and FG as congruent]"⁴. So, Rose had the conceptual understanding to use the definition of midpoint to deduce Segments EF and FG are congruent in her oral planning, but she did not explicitly convey this deduction in her written proof.

Another noteworthy set of FG1 gaps occurred when students drew conclusions about equality and/or congruence. For example, in step 9 of Jim's formal proof for Problem F (see Figure 2), Jim wrote that $FM=EM$ by CPCTC. However, CPCTC's conclusion is stated for congruence, not equality. Therefore, to be axiomatically correct Jim would have to split his step 9 into two deductions (i.e., $\overline{FM} \cong \overline{EM}$ by CPCTC and then $FM=EM$ by definition of congruence). This is not a notational error of miswriting $FM=EM$ for $\overline{FM} \cong \overline{EM}$ (i.e., FG3 gap) because Jim orally stated it as "FM equals EM" not clarifying "FM or EM" as lengths or segments. However, there are two possibilities in Jim's curtailment. One is that Jim combined steps because he thought that congruence implied measurement equality. A second possibility is that Jim confounded the concepts of equality of lengths and congruence of corresponding segments, leading to an *implicit* curtailment of steps involving these two ideas and thus a FG1 gap of length 2 (we return to this idea later).

Problem F:

Given: Point M is the midpoint of \overline{AB} and \overline{CD}

Prove: $EM = FM$



Jim's Formal Proof

- | | |
|---|---------------------|
| 1) Point M is the Midpoint of \overline{AB} and \overline{CD} | 1) Given |
| 2) $\overline{AM} \cong \overline{MB}$ | 2) Def. of Midpoint |
| 3) $\overline{CM} \cong \overline{MD}$ | 3) Def. of Midpoint |
| 4) $\angle DMB \cong \angle CMA$ | 4) Vertical Angles |
| 5) $\triangle DMB \cong \triangle CMA$ | 5) SAS |
| 6) $\angle DBM \cong \angle CAM$ | 6) CPCTC |
| 7) $\angle FMB \cong \angle EMA$ | 7) Vertical Angles |
| 8) $\triangle FMB \cong \triangle EMA$ | 8) ASA |
| 9) $FM = EM$ | 9) CPCTC |

Figure 2. Jim's demarcations of the given diagram and formal proof for Problem F

Confounding congruence of spatial objects and equality of object measures was common among students. Another example, with angles instead of segments, occurs in step 5 of Jenny's written proof for Problem G (see Figure 3). Because Jenny stated in her oral planning "I am going to mark in the givens that these are perpendicular [marks in right angles at Angles PTQ and RWU on the given diagram]," it is clear that she understands that the intersection of perpendicular segments forms right angles. But in her written proof she wrote " $\angle RWU \cong \angle PTQ \cong 90^\circ$ by definition of perpendicular". However, the definition of perpendicular segments on the reference sheet states that "two lines (or segments) that intersect form a right angle," which does not include that intersecting angles are congruent or that their measure is 90 degrees. To be axiomatically correct, Jenny would need to write

⁴ Also note that Rose confounds equal length and congruence.

three separate deductions⁵: 1) Angles RWU and PTQ are right angles by definition of perpendicular segments, 2) Angles RWU and PTQ are congruent by All Right Angles are Congruent Theorem, and 3) $m\angle RWU = m\angle PTQ = 90^\circ$ by definition of right angles. Like Jim with length, Jenny confounds the concepts of angle measure equality and congruence, leading to an implicit curtailment of steps involving these two ideas and thus a FG1 gap of length 3.

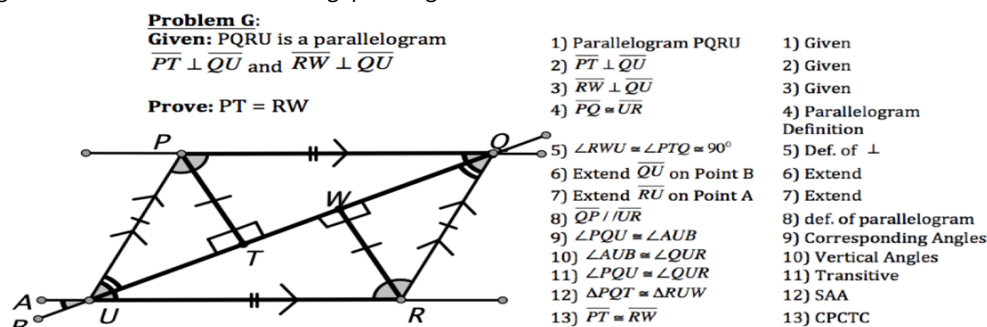


Figure 3. Jenny's demarcations on the given diagram and formal proof for Problem G

Later, in step 13, Jenny concluded her formal proof by stating that Segments PT and RW were congruent by CPCTC. However, Problem G asked her to prove that length PT equals length RW. So, to complete her formal proof, she needed to write another deduction stating that " $PT=RW$ by definition of congruence segments". The interviewer noticed that Jenny had only proven congruence of segments and not equality of lengths and asked Jenny if she had proven what the problem had asked to her prove. Jenny replied "yeah", but did notice that the statement she needed to prove was equality and not congruence. Jenny addressed this difference by stating, "Well, I think it goes without saying that they are the same messages getting across," providing evidence that Jenny confounded congruence and equality leading to an implicit curtailment of formal steps. Since Jenny is missing one deduction at the end of her proof this is an instance of FG1 of length 1, a gap that many students did throughout the interviews. Another common example of this confounding was when students wrote deductions like " $\angle 1 + \angle 2 \cong \angle 3$ by Angle Addition Postulate". On the reference sheet, the Angle Addition Postulate is stated for equality of angle measurements, not congruence, which is the case for many high school geometry textbooks.

Formalization gap 2 (FG2)–Incorrect justification

Formalization gap 2 (FG2)–Incorrect justification occurred when students incorrectly cited a theorem or definition in the justification for a valid deductive step, either by stating a closely related theorem/property or providing a justification that was axiomatically disconnected. In these cases, the students orally stated and/or gestured to something during their oral planning or explanations that provided evidence that they intuitively understood the correct justification/theorem for the step, however, they just misnamed the theorem being applied. For example, two students correctly gestured to Alternate Interior Angles in a parallel lines configuration and orally stated this pair of angles were congruent, but wrote the justification as Alternate Exterior Angles Theorem in their formal proof. Another example of the FG2 gap is when students cited the definition of isosceles triangles to justify that base angles are congruent when the correct justification is Base Angles Theorem.

Of particular note was when the students wrote a theorem's name when the correct justification was its converse. This error can be classified as formalization gap or fatal logical gap depending on the specific context of how it occurred during the interviews. If both the theorem and its converse were true (e.g., AIA and AIA converse) in the axiomatic system (i.e., the reference sheet) and the student interchanged them, then it was misnaming and was coded as a FG2 gap. If the converse is not true or not yet proven in the axiomatic system and the student uses it as justification, then it is a fatal logical gap (see LG2 gap section below).

We define *justification that is axiomatically disconnected* as a justification in which it was not clear from what the students said in their oral explanations and what they wrote in their formal proof how the stated (orally and written) justification connected to the axiomatic system (i.e., the reference sheet). The most common observed example of a justification that was axiomatically disconnected was when students stated that they constructed an auxiliary line or segment and stated "construction" as its justification, which is not a justification found in the axiomatic system (more on this later in the quantitative results of the two types of reasoning gaps).

An example of this formalization gap (FG2) is illustrated with Jenny's work on Problem N (see [Figure 4](#)).

During oral planning:

Jenny: ...What I just did was I saw the whole triangle as one complete unit [outlines Triangle NET with her finger]. Because I saw these two angles [points to angles 1 and 4, which are given as congruent] as base angles, therefore it is an isosceles triangle. So, NE is congruent to ET [marks to these respective segments as congruent on the diagram].

⁵ We recognize there are other ways to state the deductions correctly such as: 1) Angles RWU and PTQ are right angles by definition of perpendicular segments, 2) $m\angle RWU = m\angle PTQ = 90^\circ$ by definition of right angles, and 3) Angles RWU and PTQ are congruent by definition of congruence. Either way it is three steps.

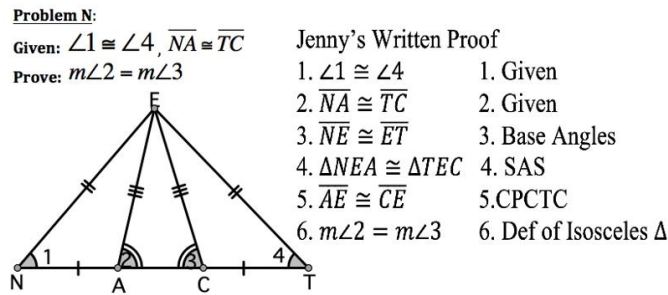


Figure 4. Jenny's demarcations on the given diagram and written proof for Problem N

Later just after writing step 3 in her proof:

IN: And what are base angles? What do you know about base angles?

Jenny: That if the base angles of a triangle are congruent, then the triangle is an isosceles. So therefore these [points to Segments NE and ET] would be the same. So, we kind of just jumped that with if the base angles [points to angles 1 and 4] are the same, then legs [outlines Segments NE and ET with the back of her marker] of the triangle are the same.

In Jenny's oral planning, she stated "Because I saw these two angles [points to angles 1 and 4] as base angles, therefore it is an isosceles triangle," and similarly, as she was writing step 3 of her formal proof, she orally stated "If the base angles of a triangle are congruent, then the triangle is an isosceles." In both cases, Jenny's oral statements are citations of the Converse of Base Angles Theorem, not the Base Angles Theorem, which she had cited as her justification for step 3 in her formal proof. So, it is a FG2 gap. We discuss other examples of FG2 gaps and possible causes for this type of behavior later in the Discussion section. (It should be noted, that all students were allowed to look at the reference sheet which included the Converse of Base Angles Theorem. However, many students, like Jenny, did not look at it unless they were struggling to come up with a name for theorem, which in this case Jenny did not seem to need to use it.)

Formalization gap 3 (FG3)–Clerical

Formalization gap 3 (FG3)–Clerical occurred when students misnamed a figure like an angle or a triangle in a written conclusion (i.e., notation error). For example, a student correctly demarcated on the diagram or gestured to an angle in the diagram (e.g., angle ABC), but wrote the angle's name in the wrong order where the vertex of the angle was not the middle letter (e.g., angle ACB).

Fatal Logical Gaps

A fatal logical gap [or logical breach] in a student's OVERALL proof reasoning occurs when the student draws a conclusion based on faulty or inadequate logic both in their oral planning/explanations and their written proof. We coded an instance as a fatal logical gap in the data if it fit one of the following two categories.

Logical gap 1 (LG1)–False conclusion

Logical gap 1 (LG1)–False conclusion occurred when students stated deductions in their oral explanations and/or in their written proofs that are false or invalid in the axiomatic system.

We illustrate a LG1 gap with Ellie's work for Problem E (see Figure 5). Prior to the beginning of the transcript, in her oral planning, Ellie deduced Segments BE and DE are congruent as well as Segments AE and ED are congruent by definition of bisects and demarcated it on her diagram. When she drew her diagram, she stated that she was drawing a square, although the quadrilateral she drew visually appeared to be more like a typical rectangle than a square (see Figure 5). It is important to note that Ellie initial struggled to recall the definition of parallelogram and only determined its definition after looking at the reference sheet, which states a parallelogram is "a quadrilateral with opposite sides parallel". She demarcated the opposite sides were parallel on her diagram, but recognized that this conclusion was something she still needed to prove.

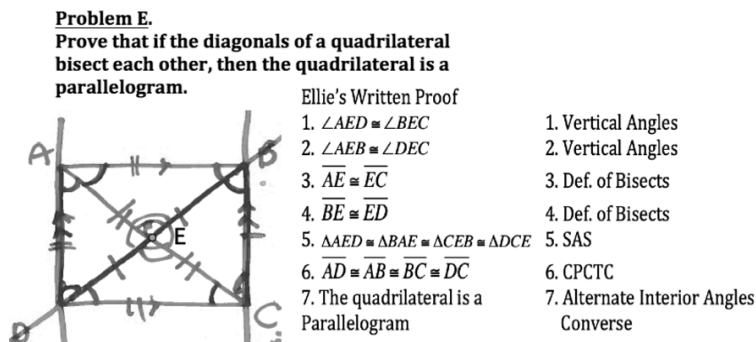


Figure 5. Ellie's drawn diagram and written proof for Problem E

During oral planning:

Ellie: Well, these would be equal angles because they are Vertical Angles [marks Angle AED congruent to Angle BEC on the diagram]. And then these two angles are Vertical Angles [marks Angle AEB is congruent to Angle CED on the diagram] too. *I think that all of these triangles would be equal by Side-Angle-Side.*

IN: Which triangles are going to be equal? Is it all four of them?

Ellie: Yeah. So, I don't know what that does for this whole thing though because these are equal [points to the four triangles AED, DEC, CEB, and BEA] I don't know what that means.

IN: But you have said that you think that all four of these are equal.

Ellie: Yeah.

IN: These inner triangles.

Ellie: Yeah.

IN: What was the reason again?

Ellie: Side-Angle-Side.

[Later in the episode during her oral planning]:

Ellie: So, like if these angles um since I said that all the triangles were congruent by Side-Angle-Side, then all of the angles would be congruent [marks the Angles DAE, BAE, ABE, CBE, BCE, DCE, CDE, ADE as congruent. See her diagram in **Figure 5**].

[Later in the episode as Ellie is writing her proof]:

Ellie: ...I can prove that it is a parallelogram because these are parallel because of the Alternate Interior Angles converse. [Ellie writes the quadrilateral is a parallelogram as her statement for step 7 of her proof].

IN: What was the reason?

Ellie: Because of the Alternate Interior Angles converse.

IN: What specifically about Alternate Interior Angles converse? So, what angles are you specifically talking about?

Ellie: BAE and BCE or um [I mean] DCE. So, should I write that?

IN: So, if I was just reading it would I be able to figure out what you meant by Alternate Interior Angles, like what angles are you saying are congruent?

Ellie: Well because I marked it [on the diagram]. ...

IN: So, you are saying that because of Alternate Interior Angles converse that now you know that it is a parallelogram. And you are saying because you marked it [points to Angles BAE and DCE on her diagram] it should be clear.

Ellie: Yeah.

In her oral planning, Ellie begins by correctly deducing that angles AED and BEC as well as angles AEB and DEC are congruent by Vertical Angles Theorem. In her oral planning and shown in step 5 of her formal proof, Ellie appeared to use steps 1-4 as premises to deduce that all four triangles (i.e., triangles AED, BAE, CEB, and DCE) were congruent by the SAS Postulate. This deduction is invalid because it is only true if the rectangle is a square, so this is a *fatal logical gap (LG1)* in Ellie's proof reasoning. One possibility is that the breakdown in Ellie's reasoning is due to an incomplete spatial structuring of her diagram, which must include an appropriate articulation of corresponding parts of triangles.

Later in her oral planning, Ellie stated that since all the small triangles were congruent (i.e., the invalid deduction in step 5), she could conclude that all the remaining angles of the four congruent triangles (i.e., angles DAE, BAE, ABE, CBE, BCE, DCE, CDE, ADE) are congruent, another fatal logical gap (LG1) since this conclusion is also invalid. This is an additional invalid deduction because even if she used as a premise her earlier invalid deduction that the four interior triangles were congruent, she could only deduce by CPCTC that corresponding angles of the four triangles were congruent ($\angle ABE \cong \angle CDE \cong \angle CBE \cong \angle ADE$ and $\angle BAE \cong \angle DCE \cong \angle BCE \cong \angle DAE$)—not all eight angles are congruent to each other. It is only in Ellie's oral planning that she stated this invalid deduction because she never wrote it in her formal proof. This gap was likely created by an incorrect spatial structuring of diagram; Ellie orally stated that she drew a square. And even though she did not draw a square, she may have geometrically structured her diagram with the properties of a square, which were not given with the problem.

Ellie's incorrect spatial structuring might also be connected to the fact that Problem E does not provide a diagram nor explicitly state what is given and what needs to be proved. Herbst and Brach (2006) found that geometry students said that proofs of theorems without explicit diagrams-givens-prove format would either be done by the teacher or as part of a class discussion. This type of task would almost never have given to students to do individually or, if it was, the students saw it as the responsibility of the teacher to provide them with hints/details as to what was given and to what needed to be proved (Herbst & Brach, 2006). However, we argue that in order to get deeper understanding about students' overall proof reasoning, we need to investigate how students construct proof problems presented in different formats. Also, proof problems like Problem E in the theorem or conditional statement format were found in the students' textbook.

Logical gap 2 (LG2)–Direct or indirect circular logic

Logical gap 2 (LG2)–Direct and indirect circular logic occurred when the student either used the theorem/conjecture that they were trying to prove in their proof of that theorem (direct circular), or they justified a deduction using a theorem which is true, but had not yet been proven in the axiomatic system (indirect circular). As an example of indirect circular reasoning, suppose the student is asked to prove Theorem A, "opposite sides of parallelogram are congruent" and the student uses Theorem B, "diagonals of a parallelogram bisect each other", as justification in their proof of Theorem A. Although, Theorem B is true, it has yet to be proven by the student, his teacher, or his class. Later, when the student is asked to prove Theorem B, he used Theorem A in the justification. Since Theorems A and B are used to prove each other this student would be exhibiting an indirect circular reasoning error. We illustrate this category using Rose's work for Problem O. The episode picks up at the beginning of episode in which Rose reads the problem aloud followed by part of her oral planning for the proof.

Rose: Problem O. Prove that opposite sides of a rhombus are parallel. Well that is easy I already proved it.

IN: How did you do that?

Rose: A rhombus is a parallelogram.

IN: Okay... Now can you prove that a rhombus is always a parallelogram?

Rose: It is in the definition.

IN: Look and see what the definition says [points to the reference sheet].

Rose: [Looks at the reference sheet] A quadrilateral with four congruent sides.

IN: Okay... Can you prove to me that rhombus is a parallelogram?... How do you know that? ... Could you write something out formally that says [proves] that? Like a formal proof.

Rose: [Writes out the initial proof shown in [Figure 6](#)] There you go.

1) rhombus
2) opposite sides of a rhombus are //

1) given
2) def rhombus QED

Figure 6. Rose initial proof for Problem O

IN: ... So how do we know that a rhombus is a parallelogram?

Rose: [Laughs]. Because its sides are parallel.

IN: How do you know that the sides are parallel? Because in the book we used it was a quadrilateral with four congruent sides.

Rose: Okay. ... [Rose first draws a diagram of a Rhombus ABCD and then draws in its diagonals labeling their intersection point as Point E]. That the diagonals like bisect each other. Oh, I don't really need it. [Rose marks angles CED and AEB are congruent on the diagram]. So, vertical angles. [Long pause]. Yeah, I actually do need another side. No, I can't prove it that way. Can I use that the diagonal bisects each other?

IN: Do you know that the diagonals bisect each other for a rhombus?

Rose: Yes. If it is true for a parallelogram it would have to be true for a [rhombus]. [Rose marks that Segment CE is congruent to Segment AE as well as Segment BE is congruent to Segment DE on the diagram]. Now they are congruent because of Side-Angle-Side. And because of the CPCTC these angles are congruent [marks Angles DCA and BAC as congruent on the diagram]. And they are alternate interior angles, which means that the sides have to be parallel.

After reading the problem, Rose immediately stated that a rhombus is a parallelogram as a part of the definition of a rhombus, and therefore “opposite sides of rhombus are parallel.” This suggests that Rose was operating with the understanding that a rhombus is always a parallelogram, a property that Problem O was in essence proving, a LG2 direct circular reasoning gap (Figure 7). In response, the interviewer asked Rose to read the definition of rhombus on the provided reference sheet (i.e., axiomatic system), which only stated “a quadrilateral with all congruent sides”. When asked by the interviewer “How do you know that a rhombus is parallelogram?” Rose replied “Because its sides are parallel.” It is only after the interviewer stated that this notion was what she was being asked to prove in the problem, that Rose abandoned the notion that rhombuses had opposite sides parallel as part of the definition of rhombus. Rose then orally stated that $CE=EA$ and $BE=ED$ because the diagonals of a rhombus bisect each other. When asked how she knew this, Rose stated “If it is true for a parallelogram it would have to be true for a [rhombus].” The fact that she stated that the diagonals of a parallelogram bisect each other guarantees the same is true for a rhombus is based on the unproven premise that all rhombuses are parallelograms. Since she used the statement that she is trying to prove in her proof of the statement, Rose made a direct circular reasoning error, a LG2 gap⁶. We propose a theoretical explanation of this behavior in the discussion section of the article.

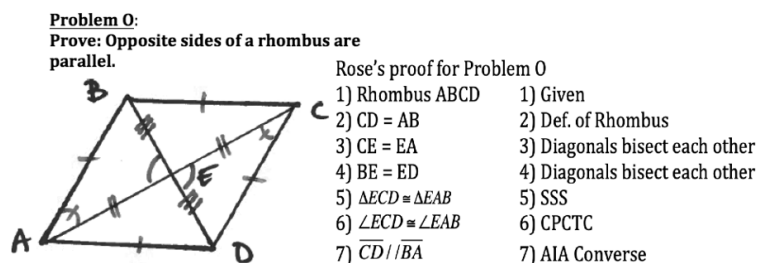


Figure 7. Rose’s drawn diagram and written proof for Problem O

Quantitative Results of the Two Types of Reasoning Gaps

In our summary quantitative analysis, we only classified and counted instances of formalization gaps and fatal logical gaps when the reasoning gaps appeared in both students’ formal/written proofs and their oral explanations/plans. The seven students each completed 12 proofs, creating a total of 84 problems for analysis. Table 1 shows the breakdown of the formalization and fatal logical gaps according to their categories. Table 2 and Table 3 shows the breakdown per student and per problem of the formalization gaps and fatal logical gaps, respectively. We found that there were 156 instances of formalization gaps while only 9 instances of fatal logical gaps. The mean of formalization gaps per student was 22.29 and the median was 21 whereas the mean of fatal logical gaps per student was 1.29 and the median was 1. Approximately 18% of students’ formal proofs (15/84) had no gaps in them and were completely correct, and each student produced at least one proof without any gaps. Approximately 72.5% of students’ formal proofs (61/84) had only formalization gaps in them compared to the 9.5% of students’ formal proofs (8/84) that had at least one fatal logical gap. FG1 Curtailment gaps accounted for approximately 49% of the reasoning gaps found in the data. Overall, the data suggests that most of the participating students had sound intuitively correct proof reasoning and logic, but they lacked understanding of the details needed to stand up to axiomatic scrutiny for formal proofs.

Table 1. Number of formalization gaps and fatal logical gaps separated by category

Formalization gaps (FG) categories	# of instances	Fatal logical gaps (LG) categories	# of instances
Formalization gap category 1 (FG1)–Curtailment	81	Fatal logical gap category 1 (LG1)–False conclusion	5
Formalization gap category 2 (FG2)–Incorrect justification	65	Fatal logical gap category 2 (LG2)–Direct or Indirect circular logic	4
Formalization gap category 3 (FG3)–Clerical	10		
Total formalization gaps	156	Total fatal logical gaps	9

Table 2. Formalization gaps (FG) breakdown per student and per problem

	Rose	Jenny	Ellie	Amy	Jim	Tinny	Sarah	Total FGs per problem	Average of FG per problem
Problem S	1	0	1	0	0	1	1	4	0.57
Problem A	0	0	3	0	1	3	0	7	1.00
Problem M	1	1	1	1	2	2	1	9	1.29
Problem L	3	0	0	0	0	0	0	3	0.43
Problem D	2	3	2	3	2	2	2	16	2.29
Problem N	1	3	2	1	6	2	5	20	2.86
Problem E*	1	2	1	1	2	3	0	10	1.43
Problem F	2	1	0	1	2	4	1	11	1.57
Problem H*	2	3	3	4	5	3	2	22	3.14
Problem G	4	3	4	4	4	4	4	27	3.86
Problem O*	3	3	2	3	3	1	4	19	2.71
Problem I	1	1	2	1	2	0	1	8	1.14
Total per student	21	20	21	19	29	25	21	156	22.29

Note. *Problems that are in the theorem or conditional statement format and no diagrams were provided

⁶ Any time a gap occurred in multiple deductive steps that could have been written as one deductive step, like Rose’s steps 3 and 4 in Problem O, we only classified it as one gap.

Table 3. Fatal logical gaps (LG) breakdown per student and per problem

	Rose	Jenny	Ellie	Amy	Jim	Tinny	Sarah	Total FGs per problem	Average of FG per problem
Problem S	0	0	0	0	0	0	0	0	0
Problem A	0	0	0	0	0	0	0	0	0
Problem M	0	0	0	0	0	1	1	2	0.29
Problem L	0	0	0	0	0	0	0	0	0
Problem D	0	0	0	0	0	0	0	0	0
Problem N	0	0	0	0	0	0	0	0	0
Problem E*	0	0	2	0	0	0	0	2	0.29
Problem F	0	0	1	0	0	0	0	1	0.14
Problem H*	0	0	0	0	0	0	0	0	0
Problem G	0	0	0	0	0	0	0	0	0
Problem O*	1	0	0	0	1	0	1	3	0.43
Problem I	0	0	1	0	0	0	0	1	0.14
Total per student	1	0	4	0	1	1	2	9	1.29

Note. *Problems that are in the theorem or conditional statement format and no diagrams were provided

DISCUSSION

Dealing with the Givens

We did not include in our gap count when students did not write the givens in their formal proofs. For reference, we found that there were 27 instances in which the students did not explicitly write the givens in their formal proofs. Most of these instances (23) were done by the two students (i.e., Ellie and Tinny) who were in the normal track geometry class. Both teachers of the participants stated that they required students to write the givens in their proofs. The reason for not including it in our gap count is that it is debatable whether students should or should not be required to write the givens in their formal proofs and whether that constitutes a formalization gap. On one hand, a function of writing the givens in a proof is to explicitly state the givens that the individual is actually using as premises (i.e., relevant) to implement later deductions a proof. On the other hand, it can be argued when the givens are explicitly stated in the problem that writing the givens is unnecessary. In fact, nine of the proof problems used the GPFT format as shown in Problem M (see [Figure 1](#)). The remaining three problems were stated in theorem format, with no diagrams. For these problems, the givens and what needed to be proven were not explicitly stated and had to be deciphered by the students from the provided problem statement. McCrone and Martin (2004) found that 12 out of 18 students in their study could not correctly identify the given statements from a proof problem using the if-then format. For reference, in this study, there were 18 instances of not writing the givens for GPFT problems and 9 instances when the problem was presented in theorem format. So, it could be argued that stating the givens for theorem format problems is important, whereas for the GPFT problems, the students could argue that the givens are already stated. Due to this debate, we chose not to classify when students did not write the givens in their proof as a formalization gap.

Marking Givens on Diagrams

In some cases of FG1, students thought that if they clearly demarcated givens on the diagram (e.g., right angle, congruence etc.) there was no need to explicitly state it in their formal proof. There are several issues with this practice. First, students used diagram markings to represent different types of information. They not only demarcated givens and valid deductions on their diagram (not distinguishing them), but also invalid deductions that they later convinced themselves were untrue, intuitive notions that were yet to be proven, and deductions that were true but irrelevant to their proof's logical structure. Second, demarcations on diagrams provide no insight into the logical sequence in which the deductions were made nor justifications the student used to draw these conclusions. Therefore, students' reasoning in making demarcations on a diagram, which are never sufficient for stating deductions for formal proofs, generally cannot be fully understood without listening to students' oral planning and seeing their gestures and demarcations on the diagram as they do it.

Congruence of Spatial Objects Versus Equality of Their Measurements

In approximately 64% of FG1 gaps (52/81), students confounded congruence and equal measurement. Because this confounding gap could be caused (a) by an explicit curtailment in which students thought the distinction was unnecessary, or (b) students misunderstanding the difference and relationship between congruence and equal measurement, the interviewer explicitly asked every student, "What is the difference between congruence and equality or do they mean the same thing (i.e., synonyms)? The students responded in one of two ways: 1) there is no significant difference between congruence and equality (i.e., they are just two different ways in geometry of saying that items were the "same") (6 instances) or 2) they did not know of a difference between congruence and equality (1 instance). There were three problems (Problem F, Problem N, and Problem G) in which the statement that needed to be proved was measurement equality (e.g., $EM=FM$ or $m\angle 2 = m\angle 3$). Of the 21 formal proofs that students wrote for these problems, none had a correct final deductive step that correctly deduced the required measurement equality. Students either stated the measurement and equality conclusion and justified it with a theorem that was not stated for equality, like Jim did for Problem F (see [Figure 2](#)). Or students only proved the final deductive step for congruence and did not write a deduction involving measurement and equality, like Jenny did for Problem G (see [Figure 3](#)). This evidence suggests that the participants did not recognize the difference between the geometric concepts of congruence as it relates to spatial objects (e.g., segments, angles, shapes) and equality as it relates numerical values (lengths, angle measures, etc.).

Findings about FG2 and FG3 gaps

The second most common formalization gaps among the students was FG2 Incorrect Justification, which accounted for approximately 40% (65/165) of all reasoning gaps. In all instances of FG2, the students made oral explanations and gestures to the diagram that showed that they intuitively understood the correct justification, they simply misnamed it. This suggests that many students intuitively understand which claims are valid, but struggle with citing correct justifications from the axiomatic system in their written proof.

Only 6% of reasoning gaps were classified as FG3 gaps. We see some of the FG3 gaps as part of Anderson's (1989) error classification of a "slip" which is "characterized by the fact that the subject does not reliably make that error and can self-correct when the error is pointed out" (p. 344). Anderson (1989) also argued "that slips can be traced to losses from working memory of critical information for solving a problem. Thus, when memory load goes up, slips increase." (p. 344). This suggests that students might be making notational mistakes when their memory load is high which causes them to unconsciously make errors. It is also possible that some students struggle with understanding the procedure for naming angles, triangles, etc. in a congruence statement. This small percentage of FG3 gaps suggests that most students correctly named their shapes in their conclusions in their formal proofs.

Auxiliary Segments and Formalization Gaps

There were two proof problems, resulting in 14 proofs, that required students to draw auxiliary segments, with all students doing so. However, for six out of 14 instances, students did not write a deduction stating that they could construct an auxiliary segment on their diagram, thus exhibiting FG1 gaps. The other remaining eight instances, students wrote a deduction in their proof about the auxiliary segment and wrote the justification of "Construction", which is disconnected from the axiomatic system, thus exhibiting FG2 gaps. Of students who used the justification of "construction", when asked what this justification meant responded in some form with either authoritarian proof scheme (Harel & Sowder, 1998) like "that is what the teacher told me to do" or "you can do any type of construction/drawing on the diagram you want" or some combination of both. There was not a single instance in which a student provided the proper axiomatic justification, Postulate 5, from the reference sheet which stated that "Through any two points there exists exactly one line" (i.e., Euclid's first postulate in Book 1 of *Elements*). This postulate is what the students' textbook used as the correct justification for when an auxiliary line or line segment is drawn in a proof. This suggests that these students did not understand the axiomatic justification for drawing auxiliary segments, did not believe it is important to state this deduction, or perhaps their teachers used "construction" as a justification. Although, students did not provide a correct axiomatic justification for constructing an auxiliary segment in the diagram in their formal proof, they did explicitly draw it in the diagram and use the auxiliary segment to help complete their proof. This suggests that they intuitively understood the deductive step, they just did not formalize it correctly to stand up to axiomatic-based scrutiny. Thus, all instances were classified as formalization gaps.

Quantitative Findings and Analysis for Fatal Logical Gaps

There were only 5 instances of LG1 gaps and 4 instances of LG2 gaps occurring in the interviews. The two students who committed the LG1 gaps were Ellie (4 instances) and Tinny (1 instance) whom were in normal track geometry class. As was discussed with Ellie's proof for Problem E (see [Figure 5](#)), when LG1 gaps occurred, many times students were either operating with incomplete spatial structuring, like when Ellie did not appropriately articulate corresponding parts of triangles, or incorrect spatial structuring, like when Ellie seemed to be operating as though her diagram had more details (it was square) than what was actually provided (i.e., it was a quadrilateral). LG2 gaps were when students either justified a deduction using a theorem which was true, but had yet to be proven in the axiomatic system (i.e., indirect circular logic), or when they committed a direct circular logic error. We illustrated the LG2 gap with Rose's work for Problem O. When students committed a LG2 gap they seemed to be struggling with understanding what they were allowed to use as justifications from the axiomatic system. It is only from viewing both the students' formal proofs and their oral planning/explanations that we gain additional insight as to why students committed these types of fatal logical gaps.

Integration with the Theory of Learning Progressions

To further interpret our results, we use the geometry learning progression originated by van Hiele (Clements & Battista, 1992) and elaborated by Battista (2007) (see [Table 4](#)). Relating proof writing to the original van Hiele learning progression, Senk's (1989) research suggests that success in a proof-oriented high school geometry course requires students entering the course to be reasoning at least at van Hiele level 3. Senk found that less than 22% of students entering below level 3, but 57%, 85%, and 100% at levels 3, 4, and 5, respectively, mastered proof writing. Unfortunately, research suggests that over a variety of curricula in a variety of countries and using a variety of assessments, a reasonable estimate for the percent of students who achieve (on posttests) Level 2 or higher reasoning in the van Hiele Learning Progression⁷ in any of grades 5-9 is 36% (Battista, 2007, 2019; Clements & Battista, 1992). Consistent with this estimate, research has found that (a) only about 31% of high school students have achieved Level 2 reasoning before high school geometry, and (b) only about 60% of high school students achieve Level 2 reasoning by the end of high school geometry (Clements & Battista, 1992; Senk, 1989), suggesting that the majority of high school students are not operating at a level of geometric reasoning sufficient for success in a proof-oriented course. Our LP analysis below corroborates and helps explain these previous findings. Battista's (2007) elaborated learning progression for reasoning about geometric shapes ([Table 4](#)) explains several aspects of our students' proof reasoning. We, first, provide evidence of the current

⁷ The review of research data is based on the original van Hiele levels as reported in Clements and Battista (1992).

study's students operating at Levels 2.3 and 3.3, then we discuss how the current results suggest elaborations of Levels 3.3, 3.4, and 4.

Table 4. Summary of Battista's (2007) LP for reasoning about geometric shapes

Level 1: Visual-holistic reasoning
Students identify, describe, and reason about shapes according to their appearance as visual wholes.
Level 2: Analytic-componential reasoning
Students explicitly attend to, conceptualize, and specify shapes by describing their parts and spatial relationships between parts.
2.1. <i>Visual-informal componential reasoning.</i> Students describe parts and properties of shapes informally and imprecisely because they do not possess the formal conceptualizations for precise property specifications.
2.2. <i>Informal and insufficient-formal componential reasoning.</i> As students begin to acquire formal conceptualizations of spatial relationships between parts of shapes, they use a combination of informal and formal descriptions of shapes. But the formal portions of these descriptions are insufficient to completely specify shapes by their properties.
2.3. <i>Sufficient formal property-based reasoning.</i> Students exclusively use formal geometric concepts and language to describe and conceptualize shapes in ways that delineate a sufficient set of properties to specify the shapes. Students can use and formulate definitions for classes of shapes, but their definitions are not minimal because forming minimal definitions requires relating one property to another using logical reasoning (which occurs at Level 3.3). Instead, <i>students' conceptualizations and definitions for shapes include all of the visual characteristics they associate with that shape category, as described by formal geometric concepts</i> , which we call "prototypical defining properties." As an example, for most students, the prototypical defining properties of rectangles are: opposite sides congruent and parallel, and four right angles. These properties express in formal geometric terms the most visually salient spatial characteristics that students use in conceptualizing rectangles. Of course, there are other, less visually salient properties of rectangles. For instance, in rectangles, the congruent diagonals bisect each other" (Battista et al., 2018).
Level 3: Relational-inferential property-based reasoning
Students explicitly interrelate and make inferences about geometric properties of shapes, but the sophistication of students' property interrelationships varies greatly. Because one property can "signal" other properties, students can correctly specify shapes without naming all their properties, form minimal definitions, start to logically organize sets of properties, and distinguish between necessary and sufficient sets of conditions.
3.1. <i>Empirical relations.</i> Students decide empirically that if a shape has one property, it has another.
3.2. <i>Componential analysis.</i> By analyzing how shapes can be visually constructed one-component-at-a-time, students conclude that when one property occurs, another property must occur.
3.3. <i>Logical inference.</i> Students make logical inferences about properties with "locally logical reasoning" in which they string together logical deductions based on "assumed-true" propositions, that is, propositions that they accept as true based on their experience, intuition, or authority. <i>Students at this level use logic, but they do not question the starting points for their logical deductions.</i>
3.4. <i>Hierarchical shape classification based on logical inference.</i> Students use logical inference to reorganize their shape classifications into a logical hierarchy, and they can <i>logically justify their hierarchical classifications</i> . It becomes not only clear why a square is a rectangle, but a necessary part in of their reasoning. <i>Students can also understand and create definitions that list only one sufficient property such as "a rectangle is a quadrilateral having four 90° angles," understanding that they can deduce other properties from this one defining property.</i>
Level 4: Formal deductive proof
Students can understand and construct formal geometric proofs. That is, within an axiomatic system, they can produce a sequence of statements that logically justifies a conclusion as a consequence of the "givens."

Evidence of level 2.3. Reasoning

Example 1: Jim, Problem G (Figure 8):

Problem G:
Given: PQRU is a parallelogram
 $PT \perp QU$ and $RW \perp QU$
Prove: $PT = RW$

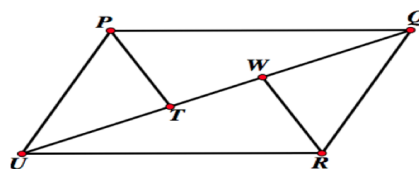


Figure 8. Problem G

Jim: [During oral planning] PQRU is a parallelogram. So that is parallel to that [points to Segments PQ and UR, respectively] and that is parallel to that [points to Segments PU and QR, respectively] ... I am thinking about how I can prove that is equal to that [points to Segments PQ and UR, respectively]. ... Oh, a parallelogram. *I think it is in the definition of parallelogram.* A parallelogram is basically a rectangle without the right angles. So that side would be equal to that side [points to Segments PQ and UR, respectively] and that side would be equal to that side [points to Segments QR and PU, respectively].

Jim does not use a minimal (formal) definition for the concept of parallelogram—a quadrilateral with opposite sides parallel—a conception that requires Level 3.4 reasoning. Instead, his definition of a parallelogram includes two prototypical defining properties—opposite sides parallel and opposite sides congruent—because he knows that both properties are true for parallelograms (see description of Level 2.3 in **Table 4**).

Example 2: Jenny, Problem N:

Jenny: [During oral planning] I saw these two angles [points to angles 1 and 4, which are given as congruent] as base angles, therefore it is an isosceles triangle. So, NE is congruent to ET.... If the base angles of a triangle are congruent, then the triangle is an isosceles. So therefore these [points to Segments NE and TE] would be the same. ...

Jenny: [Later after Jenny has written steps 1 through 5 of her formal proof.] That is like because the triangles are congruent [points to Triangles NEA and TEC] the corresponding sides [points to Segments EA and EC] are also congruent also. And then *you would have the measure of angle 2 is equal measure of angle 3 because of [pauses] I guess the definition of isosceles triangle* [writes $m\angle 2 = m\angle 3$ by definition of isosceles for step 6 of her proof]... [later, elaborating on her reasoning for step 6 in the interview]. Then they [points to Segment EA and EC] are the same. *You can jump to the base angles are the same [points to angles 2 and 3] because it is an isosceles triangle (Figure 9).*

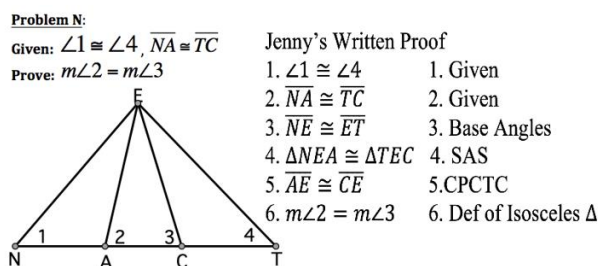


Figure 9. Jenny's proof for Problem N

There are two important points to notice about Jenny's reasoning. First is Jenny's definition of isosceles triangles, which she seemed to conceptualize as having a pair of congruent sides and corresponding base angles congruent, including two prototypical defining properties of the shape. So, she is not using the minimal definition of isosceles triangle given in the reference sheet, "a triangle with two congruent sides." Battista (2007) found that students at Level 2.3, when defining/describing a shape, almost always included all the formal properties that they associated with the shape. Other students in our study exhibited similar conceptualizations. Approximately 42% (27/65) of all FG2 gaps identified were when students wrote a definition of a shape as a justification when the correct justification was a theorem that was deduced from a formal minimal definition.

Even more, based on our current data, we hypothesize that this prototypical-defining-properties conceptualization is generally carried forward into Level 3.3, locally logical reasoning (see **Table 4**). This is illustrated by Rose's previously described reasoning for the Problem O, *prove that the sides of a rhombus are parallel*, in which she made the following statements:

"A rhombus is a parallelogram."

[Can you prove that a rhombus is always a parallelogram?] "It is in the definition."

[So how do we know that a rhombus is a parallelogram?] "Because its sides are parallel. (Citing in writing the definition of rhombuses.)"

[Do you know that the diagonals bisect each other for a rhombus?] "Yes. If it is true for a parallelogram it would have to be true for a [rhombus]."

Rose's reasoning demonstrates that, placed in an axiomatic context, she was struggling to move past her Level 2.3 conception of rhombuses that included the properties: all four sides congruent, opposite sides parallel, a special type of parallelogram, and bisecting diagonals. This conception caused her to struggle with accepting and using a minimal definition of rhombuses as quadrilaterals with all sides congruent, the definition given in the axiomatic context in which she was attempting to reason. We believe that, in this axiomatic instructional context, Rose was attempting to operate at Level 4 while still transitioning from Level 2.3 to Level 3.3, showing evidence of some reasoning at all three levels—2.3, 3.3, and 3.4, while, because of her curriculum, skipping levels 3.1 and 3.2.

Evidence of level 3.3. "Locally logical" reasoning

The distinction between "locally logical" reasoning and formal proofs is critical to understanding students' learning progression for developing proof reasoning (Battista, 2007). At Level 3.3, students' reasoning is locally logical in that they start their sequence of valid logical deductions with propositions that they accept as true based on their experience, intuition, or authority. Their reasoning is logical "but they do not question the starting points for their logical analyses." (p. 853). In contrast, at Level 4, students can understand and construct formal geometric proofs within an axiomatic system. Even though proof-oriented high school geometry courses purportedly exist in the context of an axiomatic system, it is unlikely that students view their proving processes in an axiomatic context, and instead, operate in a locally logical context in which they believe that an unordered set of axioms and theorems can be assumed true and used when needed. Furthermore, in contrast to the logical rigor of formal proofs, locally logical proofs are more informal and tend to skip steps, especially formal justifications for statements, and, as we saw above, they use definitions of shapes that include all the prototypical-defining properties students associate with the shape. In the current study, many of students' oral statements in their proof plans and written-proof oral explanations suggest

Level 3.3 locally logical reasoning or a transition from this reasoning toward Level 3.4 reasoning. Additional evidence for this type of reasoning comes from students' use of indirect circular reasoning when students use in a proof a valid theorem that has not yet been proved in the axiomatic system. In this case, students seem to use no explicit positioning in the axiomatic system—they cite what they believe is true rather than relying on axiomatic structure.

Learning progression implications for instruction

On geometric definitions: Our data suggests that our high school geometry students' conceptualizations of shape definitions are likely to include all the prototypical-defining-properties that the students know—as in Level 2.3, but continuing into Level 3.3—unless and until specific instructional attention addresses the nature of definitions (c.f., Cirillo & Hummer, 2019). In contrast to traditional instruction, in LP-based instruction, attention to the nature of definitions would occur only after the students reach Level 3.3 and teachers are trying to guide students to move into Level 3.4 (Battista, 2012; Borrow, 2000). In contrast, and inconsistent with LP-based instruction, traditional proof-oriented instruction addresses the nature of definitions at the beginning of the course, at a time when most students are not even in Level 2.3 and thus not prepared for the Level 3.4 reasoning required to make sense of such formal definitions. And students must achieve Level 3.3 and 3.4 before attempting to operate at Level 4, which is required in a proof-oriented geometry course.

On empirical exploration in proof-oriented geometry courses: Inspection of Level 3 sublevels 3.1 and 3.2 (**Table 4**) suggests that student empirical exploration, and inter-student discussion, should play an important role in students achieving first, Level 3.3, then Level 3.4, because such experiences can establish students' belief in the validity of accepted theorems and help them trust the deductions they start making in Level 3.3, double-checking deductions with empirical exploration (Battista, 2007; Borrow, 2000). We can connect this to some of Mariotti's (2006) claim that the discrepancy between empirical verification and deductive reasoning is recognized as a source of student difficulties in constructing proofs, a transition clearly demarcated in Level 3 sublevels. Furthermore, according to Mariotti (2006), Duval recognized a cognitive rupture between argumentation and proof, with argumentation consisting of the rhetoric employed to convince somebody of the truth or falsehood of a statement, and proof consisting of the logical sequence of deductions that imply the theoretical validity of a statement. Perhaps locally logical reasoning is the form of argumentation students use as they move toward structural cognitive unity, which is the relationship between the structure of an argumentation and the structure of a corresponding proof.

Elaborating level 4

The results of the current study also suggest that Level 4 in Battista's (2007, 2012) LP be elaborated into *at least* two sublevels.

Level 4: Formal deductive proof:

ORIGINAL: Students can understand and construct formal geometric proofs. That is, within an axiomatic system, they can produce a sequence of statements that logically justifies a conclusion as a consequence of the "givens."

NEW Level 4.1: When attempting to work within an axiomatic system, students give valid proofs, but with several formalization gaps in which they condense steps or ignore formal axiomatic conceptual subtleties. In some sense, this is continuation of the locally logical reasoning of Level 3.3 into Level 3.4, but now students are attempting to make their proofs explicitly formal (most often using the GPFT template).

NEW Level 4.2: When attempting to work within an axiomatic system, students give valid proofs without formalization gaps. They recognize and utilize differences among undefined terms, definitions, axioms, and theorems, and they understand *global* axiomatic sequencing and structuring in the sense that they see why, for example, theorems numbered greater than N cannot be used in the proof of Theorem N (because it often, but not always, leads to indirect circular reasoning).

Conclusions

The findings from the task-based interviews suggest that even though most of the proofs that students wrote were not formally (axiomatically) correct, most of the students used otherwise sound reasoning in most of their oral proof plans and explanations. This finding suggests that these students do not fully understand or see the need for all the details needed to write formal proofs with correct logical and axiomatic structuring—and that, most often, students struggled with formalizations rather than logical deductive arguments. Importantly, the nature of students' struggles with proof reasoning would not have been revealed if only their written proofs were examined (e.g., Ellie and Rose). Analyzing all the students' oral plans and explanations in addition to their written proofs enabled us to identify different types of gaps in students' proofs, providing deeper insights into their proof reasoning.

Most of the time in instructional and research settings, teachers and researchers evaluate students' proof reasoning only by examining their written proofs. The present study indicates that this practice is not a reliable way to genuinely understand their overall proof reasoning. A two-column proof provides limited amount of information about what students were thinking as they constructed their proofs. For example, when students are missing deductions in their written proofs, the teacher or researcher usually cannot distinguish between students who did so intentionally because they thought the deductions were obvious (e.g., Rose in Problem M) from students who did it because of some misconception or error in their reasoning (e.g., Ellie in Problem E).

In contrast, as this article illustrates, evaluating students' proof reasoning using both their oral planning/explanations and their written proofs can provide deeper insight into the students' thinking. Research studies have found that many experts quickly devise a proof-plan prior to writing their formal proofs and that these plans omit many steps (Cirillo & Hummer, 2021; Koedinger & Anderson, 1990). When asked to first orally plan their proofs, we found students, like experts, developed proof plans that were most often logically sound, but missing formal details. The major difference between experts and our students, is that the experts are able to fill in all the missing details whereas most of our students were not. This suggests that students need additional

instruction to help them translate their sound informal ideas in planning to more formal written deductions that stand up to axiomatic-based scrutiny. The additional insight gained from analyzing students as they orally plan their proofs and engage in think-alouds as they write proofs not only helps teachers and researchers better understand students' proof struggles so they can develop more effective curriculum materials and instructional interventions.

Our finding, that students struggled to understand and provide the amount of rigor and level of detail needed for valid two-column proofs, is consistent with Soto-Johnson and Fuller's (2012) and Stylianides' (2019) work in which they found that students struggled to write coherent formal proofs, but evidenced sound oral arguments. Because of the order in which the proofs were constructed, written-then-oral in Stylianides (2019), one might argue it was the order that produced the better performance on oral proofs (even though that study used group work, rather than individuals like this study). In contrast, we found that when students constructed their oral plans first then wrote their proofs, students' oral comments were less likely to skip steps. This suggests that order of oral and written proofs might not be a factor in students' performances with proofs.

Our findings also suggest that our students struggled to understand the difference between congruence and equality and the appropriate use of these concepts in their deductions in formal proofs. This is consistent with Anderson's (1989) finding with his research with the Geometry Tutor software in which almost all students did not see the difference between congruence and equality and skipped steps when the definition of congruence was needed to switch between equality of measurements to congruence of segments or angles (and vice versa). One difference between our finding and Anderson's is that we saw students not only skip steps involving congruence and equality, but also used incorrect notation in their written conclusions (e.g., $\angle A \cong \angle B \cong 90^\circ$ and $\overline{AD} + \overline{BC} \cong \overline{EF}$). We attribute this difference to asking our students to write their proofs on paper instead of using a computer software which had a selection of conclusions provided from which students could choose from. Most high school geometry textbooks seem to make a distinction between these two concepts, including our students' textbook. However, our students seem to think they are interchangeable and they are synonyms for meaning the "same as." Faulker et al. (2016) argued that students might have the following questions if the congruence and equality concepts are viewed as interchangeable "If equal means 'same as' then what does congruence mean and why do I need another vocabulary word that means the 'same as'?" (p. 15). Anderson (1989) also argued for importance of this distinction as follows:

The reader may well consider why this distinction between equality of measure and congruence is being enforced in the geometry curriculum... Congruence means that objects will be the same after rotation, reflection, and translation. Equality means the identity of two numerical measures. The distinction becomes important in the later chapters where two objects (e.g., triangles) can be equal in some measure (e.g., area) but not congruent (p. 360).

So, the question becomes how important is the distinction between congruence and equality in high school geometry? On one hand, maybe the difference between congruence and equality is too fine of a distinction so that it is unnecessary for high school students to make it. In fact, there are some older high school geometry textbooks in which the authors do not distinguish between congruence and equality, rather just use equality for triangle congruence. For instance, noted Harvard mathematician George Birkhoff⁸ and his coauthor state in their high school geometry text, "it is advisable in an elementary course to slur over or ignore some of the subtler mathematical details, for these are not suitable material for the mind of the student at this juncture" (Birkhoff & Beatley, 1940, p. 4). As an example, consider their statement of Assumption 1: "If in two triangles two sides and the included angle of one are equal respectively to two sides and included angle of the other, the two triangles are equal," noting that "other books on geometry often refer to equal triangles as 'congruent' triangles" (p. 59).

On the other hand, if we consider this distinction important, then it seems critical that teachers explicitly provide instruction to help students understand the difference between the two concepts and how to appropriately use them in formal proofs. Faulker et al. (2016) argued that teachers need to help children develop a more precise definition of equality that goes beyond "same as" in order to avoid confusion when they are introduced the concept of congruence later in geometry. They also argued that the concept of equality should be tweaked to "same in value" to explicitly state the attribute that is the same, which in many cases in geometry is a numerical value representing a measurement. In either case, more research and discussion among teachers and mathematics educators is needed to sort out whether the distinction between congruence and equality is an important one to emphasize in high school geometry.

Limitations

One limitation of our study was the small number of students ($n=7$) who participated in the interviews. So further research is needed with a larger sample of students to see if the findings generalize. Another limitation is that to study students' geometry proof reasoning in depth, the present study intentionally limited the presented proof problems to ones that use triangle congruence postulates in US high school geometry. Although the proofs covered in the present study represent a major category of proofs, and the reasoning students used in these proofs seems to represent many generalized proving behaviors, extending the current study's results to other kinds of proof problems (e.g., similarity, transformations, circles, etc.) cannot be accomplished without further research.

Another limitation is the finding of students not writing givens in their proofs. What is clear from the data is that students who did not write the givens did not seem to have a proper understanding of the function of including givens in their proofs. A few students cited the givens in their proof right before they used them as premises to draw subsequent deductions, thus exhibiting

⁸ It is worth noting that Birkhoff created a set of four postulates of Euclidean plane geometry that build on the postulates for the set of real numbers, thereby simplifying the axioms for Euclidean geometry. These axioms were later modified by the School Mathematics Study Group as a new standard for high school geometry.

an explicit understanding of the function of including the givens in their written proofs. However, of the other students who wrote the givens as their first steps in their proofs, it is unclear from the data whether these students truly understood the function of writing the givens or whether they did this step procedurally due to some established sociomathematical norm (Yackel & Cobb, 1996) in their geometry classrooms. Additional data is needed to determine which students understood the function of including givens which students did so by rote. This data would help distinguish whether not writing the givens in formal proofs should be classified as a formalization gap or not. Interconnected with this issue, to complete most of the proof problems presented, the students typically had to use all the provided givens. Future research needs to investigate situations in which students are provided more given information than is needed to complete the proof. This would not only provide insight into how students handle irrelevant proof givens it would help us better understand if students truly understand the function of writing the givens or if they are just implementing a sociomathematical norm in which they simply rewrite the givens in the first few steps.

Finally, the present study investigated in detail students' understanding of how to construct a particular type of formal geometric proof. It did not, however, attempt to relate students' ability to understand and write proofs to the purpose of proof. For instance, how did students' understanding of their proofs relate to whether or not they believed that the statements that they proved were true or not. Given the substantial amount of research that indicates that students do not necessarily understand how formal proof is related to the validity of proved statements (Battista, 2007), it would be fruitful to interrelate these two different lines of research on geometric proof.

IMPLICATIONS

Given that the participating students struggled with proof formalization rather than forming sound logical arguments, it is important for mathematics educators to decide how important formal proofs in high school geometry are, or if locally logical arguments are sufficient. Some teachers have questioned whether formal proofs in high school are appropriate in favor of more exploratory activities in which students construct informal arguments (Boyle, 2012; Knuth, 2002). Many researchers have found that most proof problems given to students in high school geometry are more ritual-based exercises, which teach students about deductive reasoning and logic and are less about explaining why geometry works (Gilbertson et al., 2013; Herbst, 2002; McCrone & Martin, 2009). However, if mathematics educators view formal proof as important, then how it is taught needs to change so students can understand the formal details needed for valid axiomatic proof. One possibility is using Dynamic Geometry Environments (DGE) to help students see the need for formal proofs and appreciate the other functions of proof (explaining, discovery, systematization, etc.) (Battista, 2007; de Villers, 2004; Mariotti, 2006). One possible instructional technique that might help is having students explore ideas and properties of shapes using DGE, form their own conjectures from their explorations, and then try to prove or disprove their conjectures using formal proof or counterexamples.

Since most of our students' errors were formalization gaps, which were most identifiable when students were asked to orally explain their thoughts as they planned and wrote their proofs, mathematics educators need to find alternate ways to assess proofs besides only reading the students' written proofs. We recognize, however, that teachers typically do not have the time or the resources to conduct one-on-one interviews with every student. Therefore, an important instructional implication of this study is that teachers need to find ways to create opportunities for students to orally explain their thinking before constructing formal proofs in ways that teachers could observe. One possibility is that, before constructing a formal proof, students could be put in small groups to explain their plans for creating the proof (while the teacher circulated about the classroom), and students could even present their plans to the whole class. This would allow teachers to formatively assess students' proof reasoning in order to address and remedy any difficulties that arise before students take a more traditional written assessment.

Another possibility is to have students submit video recordings using their phones and along with their oral plans and written proofs upload them to website set up by their teachers. Teachers might to examine samples of students' reasoning, or they might use the video recordings only when they find missing or incorrect steps in the students' written proofs. Another possibility is for mathematics educators to create some sort of computer software in which students not only type in their deductive steps, but also orally record their reasoning for each step. Then teachers could read the written proof and click the oral recording next to each deduction to hear the students reasoning for the step.

Author contributions: All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Declaration of interest: No conflict of interest is declared by authors.

REFERENCES

- Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr, & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 343-371). Lawrence Erlbaum Associates, Inc.
- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228(4698), 456-462. <https://doi.org/10.1126/science.228.4698.456>
- Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem-solving skill. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 191-230). Lawrence Erlbaum Associates, Inc.
- Battista, M. T. (1999). Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom. *Journal for Research in Mathematics Education*, 30(4), 417-448.

- Battista, M. T. (2001). How do children learn mathematics? research and reform in mathematics education. In T. Loveless (Ed.), *The great curriculum debate: How should we teach reading and math?* (pp. 42-84). Brookings Press.
- Battista, M. T. (2007). The development of geometric and spatial thinking. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 843-908). National Council of Teachers of Mathematics.
- Battista, M. T. (2008). Development of the shape makers geometry microworld. *Research on technology and the teaching and learning of mathematics: Volume 2 cases and perspectives* (pp. 131-156).
- Battista, M. T. (2011). Conceptualizations and issues related to learning progressions, learning trajectories, and levels of sophistication. *The Mathematics Enthusiast*, 8(3), 507-570.
- Battista, M. T. (2012). *Cognition-based assessment & teaching of geometric shapes: Building on students' reasoning*. Heinemann.
- Battista, M. T. (2019). *Using learning progressions to design and assess geometry curricula* [Presentation]. NCTM Research Conference, San Diego, April 2019.
- Battista, M. T., & Clements, D. H. (1995). Geometry and proof. *Mathematics Teacher*, 88(1), 48-54.
- Battista, M. T., Frazee, L. M., & Winer, M. L. (2018) Analyzing the relation between spatial and geometric reasoning for elementary and middle school students. In K. S. Mix & M. T. Battista (Eds.), *Visualizing mathematics* (pp. 195-228). Springer.
- Beman, W. W., & Smith, D. E. (1899). *New plane and solid geometry*. Ginn.
- Birkhoff, G., & Beatley, R. (1940). *Basic geometry*. AMS Chelsea Publishing.
- Borrow, C. V. A. (2000). *An investigation of the development of 6th grade students' geometric reasoning and conceptualizations of geometric polygons in a computer microworld* [Doctoral dissertation, Kent State University].
- Boyle, J. D. (2012). *A study of prospective secondary mathematics teachers' evolving understanding of reasoning-and-proving* [Doctoral dissertation, University of Pittsburgh].
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Research Council.
- Campbell, T. G., King, S., & Zelkowski, J. (2021). Comparing middle grade students' oral and written arguments. *Research in Mathematics Education*, 23(1), 21-38. <https://doi.org/10.1080/14794802.2020.1722960>
- Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24(4), 359-387. <https://doi.org/10.1007/BF01273371>
- Chen, C., & Herbst, P. (2013). The interplay among gestures, discourse, and diagrams in students' geometrical reasoning. *Educational Studies in Mathematics*, 83(2), 285-307. <https://doi.org/10.1007/s10649-012-9454-2>
- Cirillo, M., & Herbst, P. G. (2012). Moving toward more authentic proof practices in geometry. *Mathematics Educator*, 21(2), 11-33.
- Cirillo, M., & Hummer, J. (2021). Competencies and behaviors observed when students solve geometry proof problems: An interview study with smartpen technology. *ZDM—Mathematics Education*, 53(4), 861-875. <https://doi.org/10.1007/s11858-021-01221-w>
- Cirillo, M., & Hummer, J., (2019). Addressing misconceptions in secondary geometry. *The Mathematics Teacher*, 35(6), 410-417. <https://doi.org/10.5951/mathteacher.112.6.0410>
- Cirillo, M., & May, H. (2020). Decomposing proof in secondary classrooms: A promising intervention for school geometry. In *The 14th International Congress on Mathematical Education*. Shanghai, China.
- Cirillo, M., Griffin, C., Seiwel, A., & Hummer, J. (2021). What do you believe is true? A routine for proving theorems in secondary geometry. In *Proceedings of the 43rd Meeting of the North American Chapter of the International Groups for the Psychology of Mathematics Education*. Philadelphia, PA, USA.
- Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547-589). Lawrence Erlbaum Associates, Inc.
- Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. *Handbook of research on mathematics teaching and learning*, 420-464.
- Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. National Governors Association Center for Best Practices and the Council of Chief State School Officer.
- De Villiers, M. (2004). Using dynamic geometry to expand mathematics teachers' understanding of proof. *International Journal of Mathematical Education in Science and Technology*, 35(5), 703-724. <https://doi.org/10.1080/0020739042000232556>
- Dreyfus, T., & Hadus, N. (1987). Euclid may stay—and even be taught. In M. M. Lindquist, & A. P. Shulte (Eds.), *Learning and teaching geometry, K-12* (pp. 47-58). National Council of Teachers of Mathematics.
- Duval, R. (2007). Cognitive functioning and the understanding of mathematical processes of proof. In P. Boero (Ed.), *Theorems in schools: From history, epistemology and cognition to classroom practice* (pp. 137-161). Sense Publishing.
- Faulkner, V., Walkowiak, T., Cain, C., & Lee, C. (2016). Equality matters: The critical implications of precisely defining equality. *Australian Primary Mathematics Classroom*, 21(4), 11-15.
- Gilbertson, N. J., Otten, S., Males, L. M., & Clark, D. L. (2013). Not all opportunities to prove are the same. *Mathematics Teacher*, 107(2), 138-142. <https://doi.org/10.5951/mathteacher.107.2.0138>
- Glaser, B. G. (1978). *Theoretical sensitivity*. The Sociology Press.

- Glaser, B., & Strauss, A. (1967). *The discovery grounded theory: Strategies for qualitative inquiry*. Aldine Transaction. <https://doi.org/10.1097/00006199-196807000-00014>
- Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517-545). Lawrence Erlbaum Associates, Inc.
- Hanna, G. (1989). More than formal proof. *For the Learning of Mathematics*, 9(1), 20-23.
- Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877-905). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-1465-0_27
- Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234-282). Mathematical Association. <https://doi.org/10.1090/cbmath/007/07>
- Heinze, A., Cheng, Y., Ufer, S., Lin, F., & Reiss, K. (2008). Strategies to foster students' competencies in constructing multi-steps geometric proofs: Teaching experiments in taiwan and germany. *ZDM*, 40(3), 443-453. <https://doi.org/10.1007/s11858-008-0092-1>
- Herbst, P. G. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics*, 49(3), 283-312. <https://doi.org/10.1023/A:1020264906740>
- Herbst, P., & Brach, C. (2006). Proving and doing proofs in high school geometry classes: What is it that is going on for students? *Cognition and Instruction*, 24(1), 73-122. https://doi.org/10.1207/s1532690xci2401_2
- Herbst, P., & Miyakawa, T. (2008). When, how, and why prove theorems? A methodology for studying the perspective of geometry teachers. *ZDM*, 40(3), 469-486. <https://doi.org/10.1007/s11858-008-0082-3>
- Kelley, G. D. (2013). *Approaches to proof in geometry textbooks: Comparing texts from the 1980s and 2000s* [Doctoral dissertation, University of Maryland, College Park].
- Knuth, E. J. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, 5(1), 61-88. <https://doi.org/10.1023/A:1013838713648>
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14(4), 511-550. https://doi.org/10.1207/s15516709cog1404_2
- Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery* (I. Lakatos, J. Worrall, & E. Zahar [Eds.]). Cambridge University Press. <https://doi.org/10.1017/CBO9781139171472>
- Larson, R., Boswell, L., Kanold, T. D., & Stiff, L. (2008). *Geometry*. McDougal Littell.
- Mariotti, M. A. (2006). Proof and proving in mathematics education. In A. Gutiérrez, & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173-204). Sense Publishing. https://doi.org/10.1163/9789087901127_008
- Martin, T. S., & McCrone, S. M. (2003). Classroom factors related to geometric proof construction ability. *The Mathematics Educator*, 7(1), 18-31.
- McCrone, S. M. Martin, T. S.; Dindyal, J., & Wallace, M. L. (2002). An investigation of classroom factors that influence proof construction ability. In *Proceedings of the Annual Meeting [of the] North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1-4, p. 13). Athens, GA, USA.
- McCrone, S. M. S., & Martin, T. S. (2004). Assessing high school students' understanding of geometric proof. *Canadian Journal of Math, Science & Technology Education*, 4(2), 223-242. <https://doi.org/10.1080/14926150409556607>
- McCrone, S. M., & Martin, T. S. (2010). Formal proof in high school geometry: Student perceptions of structure, validity, and purpose. *Teaching and learning proof across the grades*, 204-221. <https://doi.org/10.4324/9780203882009-12>
- McCrone, S. S., Martin, T. S., Dindyal, J., & Wallace, M. L. (2002). An investigation of classroom factors that influence proof construction ability. In *Proceedings of the Annual Meeting [of the] North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1-4, p. 13). Athens, GA, USA.
- Movshovitz-Hadar, N. (2001). Proof. In L. Grinstein, & S. I. Lipsey (Eds.), *Encyclopedia of mathematics education* (pp. 585-591). Routledge.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- National Research Council. (2001). Knowing what students know: The science and design of educational assessment. Committee on the foundations of assessment. In J. W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Board on testing and assessment, center for education. Division of behavioral and social sciences and education*. National Academies Press.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8. committee on science learning, kindergarten through eighth grade*. National Academies Press.

- Pair, J., Strachota, S., & Singh, R. (2021). The academic community's perceptions of the two-column proof. *The Mathematics Enthusiast*, 18(1), 210-238. <https://doi.org/10.54870/1551-3440.1521>
- Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of 'well-taught' mathematics courses. *Educational Psychologist*, 23(2), 145-166. https://doi.org/10.1207/s15326985ep2302_5
- Senk, S. L. (1985). How well do students write geometry proofs? *The Mathematics Teacher*, 78(6), 448-456. <https://doi.org/10.5951/MT.78.6.0448>
- Senk, S. L. (1989). Van Hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education*, 20(3), 309-321. <https://doi.org/10.2307/749519>
- Shaughnessy, J. M., & Burger, W. F. (1985). Spadework prior to deduction in geometry. *The Mathematics Teacher*, 78(6), 419-428. <https://doi.org/10.5951/MT.78.6.0419>
- Sinclair, N., Cirillo, M., & de Villiers, M. (2017). The learning and teaching of geometry. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 457-488). National Council of Teachers of Mathematics.
- Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, 4(1-2), 1-98. <https://doi.org/10.1080/15366367.2006.9678570>
- Soto-Johnson, H., & Fuller, E. (2012). Assessing proofs via oral interviews. *Investigations in Mathematics Learning*, 4(3), 1-14. <https://doi.org/10.1080/24727466.2012.11790313>
- Steffe, L., & Thompson, P. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Lawrence Erlbaum Associates, Inc.
- Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38(3), 289-321.
- Stylianides, A. J. (2019). Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation. *Review of Education*, 7(1), 156-182. <https://doi.org/10.1002/rev3.3157>
- Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). *Research on the teaching and learning of proof: Taking stock and moving forward*. National Council of Teachers of Mathematics.
- Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher*, 41(5), 147-156. <https://doi.org/10.3102/0013189X12442801>
- 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. <https://doi.org/10.1007/s10649-008-9144-2>
- Wertheimer, R. (1990). The geometry proof tutor: An "intelligent" computer-based tutor in the classroom. *The Mathematics Teacher*, 83(4), 308-317. <https://doi.org/10.5951/MT.83.4.0308>
- Wu, H. H. (1996). The role of Euclidean geometry in high school. *The Journal of Mathematical Behavior*, 15(3), 221-237. [https://doi.org/10.1016/S0732-3123\(96\)90002-4](https://doi.org/10.1016/S0732-3123(96)90002-4)
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458-477. <https://doi.org/10.2307/749877>