The 14th International Congress on Mathematical Education
Shanghai, 12th –19th July, 2020

DECOMPOSING PROOF IN SECONDARY CLASSROOMS: A PROMISING INTERVENTION FOR SCHOOL GEOMETRY

Michelle Cirillo
University of Delaware

Henry May
University of Delaware

Across three years, more than 1000 students enrolled in courses that addressed proof in secondary geometry were tested at the beginning of the school year and after completing the proof unit(s). In 4 of the 20 study teachers' classrooms, students were introduced to proof through a set of 16 specially designed lesson plans that addressed particular sub-goals of proof; while students in the control group learned from the standard curriculum. The sub-goals comprise a pedagogical framework that decomposes the teaching of proof in geometry. Study findings suggest that the experimental treatment had a positive statistically significant effect on students' achievement on the proof-focused post-test.

INTRODUCTION AND THEORETICAL FRAMWORK

In his review of the research on teaching proof in geometry, Battista (2007) posed the following questions: (a) Why do students have so much difficulty with geometric proof? (b) What components of proof are difficult for students and why? and (c) How can proof skills best be developed in students? (pp. 887-888). Ten years later, in an updated review of the research on teaching and learning proof in geometry, Sinclair, Cirillo, and deVilliers (2017) reported that since Battista's review was published, although researchers had attempted to design studies to better understand and, in some cases, address the difficulties of teaching and learning proof in geometry, these studies tended to focus on only one or a few teachers or did not provide evidence of effectiveness on a large scale. Sinclair et al. (2017) recommended that more research was needed on students' development of geometric proof skills and their understanding and beliefs about the nature of proof. This paper addresses the first component of this recommendation – research on the development of students' geometric proof skills.

In terms of students' skill development, Smith (1940) identified and analyzed "three serious learning difficulties" that students have when learning proof in geometry: (1) a lack of familiarity with geometric figures; (2) not sensing the meaning of the if-then relationship; and (3) an inadequate understanding of the meaning of proof (p. 100). Thirty-five years later, Senk (1985) detailed findings from her study of 1520 students in the U.S., in which she found that only 30% of students in a full-year geometry course that covered proof reached a 75% mastery level of proof writing. Consequently, she suggested that we must immediately look for more effective ways to teach proof in geometry. She suggested that teachers do the following: (1) pay special attention to teaching students to start a chain of reasoning; (2) place greater emphasis on the meaning of proof than we do currently; and (3) teach students how, why, and when they can transform a diagram in a proof (p. 455).

Building on this earlier research as well as tasks proposed by Cirillo and Herbst (2011) and prior work conducted by Cirillo et al. (2017), Cirillo launched the *Proof in Secondary Classrooms (PISC) Project* in 2015. The PISC Project is a five-year study that takes as its premise that if we introduce proof by first teaching students particular sub-goals of proof, then students will be more successful with constructing proofs on their own. The underlying theory of the pedagogical framework rests on Grossman and colleagues' (2009) notion of "decomposition of practice" - the breaking down of a practice into its constituent parts. The ability to decompose a practice is dependent on naming the constituent parts so that instructors can provide targeted feedback on students' efforts to enact particular components of practice. Although Grossman et al. (2009) considered decomposition in the context of teacher education, we argue here that the notion of decomposition can be applied not only to pedagogical practices but to mathematical processes as well (see Kobiela & Lehrer (2015) for a related example on the practice of defining). By decomposing complex practices, instructors can support learners first to attend to and then to enact the essential elements of a practice. In this vein, the research question posed here is: *When provided with instruction based on a pedagogical framework that decomposes proof in geometry into particular sub-goals of proof, how do students' scores on a set of proof tasks compare for students in the experimental and control groups?*

METHODS

The control group received regular instruction which ranged from a conventional geometry course focused on Euclidean geometry to a collection of proof units spread out over several years of 'integrated math." The experimental group received instruction based on the Geometry Proof Scaffold which we describe next. Data collection instruments for this aspect of the study included a pre-test and post-test. These instruments are described below.

Intervention: The Geometry Proof Scaffold (GPS) and the PISC Curriculum

The Geometry Proof Scaffold (GPS) was informed by the research literature and two related research projects described elsewhere by Cirillo and colleagues (see, e.g., 2008; 2017). Collectively, these studies provided evidence that when it comes to teaching proof in geometry: (a) the introduction to proof is particularly difficult for students and their teachers; (b) even experienced, "well-prepared" teachers may not feel confident in their strategies for introducing proof; and (c) curricula provide inadequate support for teaching proof in geometry. The GPS, a pedagogical framework organized around a set of nine sub-goals for teaching proof, was developed to address these and other well-documented challenges described in the literature. It describes the competencies that students must understand and be able to do in order to be successful with proving. The GPS simplifies the task of proving so that understanding can be built in progressive steps toward the larger goal of doing proof (i.e., developing conjectures, proving theorems, etc.). See the full GPS in Figure 1.

The Proof in Secondary Classrooms or *PISC* Curriculum is a set of 16 lesson plans and student activity sheets based on the ideas from the GPS. The goal of these lessons is to scaffold the introduction to proof by teaching particular competencies necessary for students to be able to write proofs on their own. In this sense, proof has been "decomposed," whereby students learn particular sub-goals of proof one at a time so that they are able to draw upon all of the competencies to participate in the reasoning and the discourse of proving. These lessons served as the intervention for the experimental group. A sample task for addressing the Coordinating Geometric Modalities sub-goal (i.e., translating notation into a diagram) is provided in Figure 2. A sample task for the Drawing Conclusions sub-goal (i.e., drawing a conclusion from a given statement and a definition) is provided in Figure 3.

Sub-Goals	Descriptions	Competencies
Knowing Geometric Concepts	Highlights the importance of knowing the building blocks of geometry	1) Having an accurate "mental picture" of a geometric concept (i.e., having a concept image
		2) Being able to verbally describe geometric concepts, ideally being fluent with one or more definitions of the concept (i.e., having or developing a formal concept definition)
		3) Determining examples and non-examples of a geometric concept
		4) Understanding connections between classes of geometric objects, where they overlap, and how they are contained within other classes (i.e., understanding mathematical hierarchy)
Defining	Highlights the nature of definitions, their logical structure, how they are written, and how they are used	1) Being able to write an economical definition for a mathematical object that identifies which class of objects it belongs to and what makes the object special (i.e., the specific difference)
		2) Knowing that definitions are not unique (i.e., geometric objects can have different definitions)
		3) Knowing how to write definitions as conditional and biconditional statements and understanding that definitions are "reversible"
Coordinating Geometric Modalities	Highlights the ways in which the mathematics register draws on a range of modalities	1) Translating between language and diagram
		2) Translating between diagram and symbolic notation
		3) Translating between symbolic notation and language
Conjecturing	Recognizes that conjecturing is an important part of mathematics and proving, in particular	1) Understanding that empirical reasoning can be used to develop a conjecture but that it is not sufficient proof of the conjecture
		2) Seeking out counterexamples to test conjectures and knowing that only one counterexample is needed to disprove a conjecture
		3) Being able to turn a conjecture into a testable conditional statement
		4) Understanding that when exploring the validity of a conjecture, you are testing it for every case
Working with Diagrams	Highlights the importance of reading, creating, and interacting with diagrams	1) Knowing how to read a diagram and understanding what can and cannot be assumed about a diagram
		2) As appropriate, sketching a diagram to be used in a proof, and knowing how and when to add an auxiliary line to a diagram
		3) Knowing how to mark a diagram based on what is known to be true
Drawing Conclusions	Presents the idea of open-ended tasks that lead to conclusions drawn from assumptions	1) Using axioms, postulates, definitions, and theorems (or combinations of these) to draw valid conclusions from some "Given" information
		2) Knowing when it is appropriate and how to use information "read" from a diagram to draw valid conclusions
Understanding Common Sub-Arguments	Recognizes that there are common short sequences of deductions that are used frequently in proofs and that these pieces may appear relatively unchanged from one proof to the next	1) Recognizing a sub-argument as a branch of proof and how it fits into the larger proof
		2) Understanding what valid conclusions can be drawn from "Given" information and how the conclusions can be chained together to make a sub-argument (i.e., knowing some commonly occurring sub-arguments)
		3) Understanding how to write a sub-argument using acceptable notation and language (where "acceptable" is typically determined by the teacher)
Understanding Theorems	Highlights the nature of theorems, their logical structure, how they are written, and how they are used	1) Being able to identify the hypothesis and conclusion of a conditional statement and then writing particular "Given" and "Prove" statements for the conditional statement, typically, making use of a generic figure.
		2) Understanding that a theorem is not a theorem until it has been proven
		3) Understanding that one cannot use the conclusions of the theorem itself to prove the conclusions of that theorem (i.e., avoiding circular reasoning)
		4) Understanding the connection between logic and a theorem, for example, knowing how to write the converse of a conditional statement, and knowing that theorems are only sometimes biconditionals
Understanding the Nature of Proof	Highlights the nature of proof, proof structure, and how the laws of logic are applied	1) Understanding that the only way to sanction the truth of a conjecture is through deductive proof (rather than empirical reasoning), and recognizing that once you prove that something is true for an arbitrary member of a class of geometric objects, then it is true for any member of that class
		2) Strategizing a plan for a proof by considering what is known and exploring a pathway toward the conclusion (i.e., the problem-solving aspect of proof)
		3) Understanding that proofs follow the laws of logic and are constructed using axioms, postulates, definitions, and previously proven theorems

Figure 1: Geometry Proof Scaffold: A Pedagogical Framework for Teaching Proof

(© Michelle Cirillo, University of Delaware, USA)

Sketch, label, and mark a single diagram with the given features.

- $\overline{AB}$ has E as a midpoint
- $\overline{AB} \perp \overline{CD}$, and $\overline{CD}$ intersects $\overline{AB}$ at B
- $\overline{BD} \cong \overline{BE}$
- Which line segments are congruent to each other? Justify your answer.

Figure 2: A Sample Task for the Coordinating Geometric Modalities Sub-Goal

What conclusion(s) can be drawn from the given information?

- **Given:** $\overline{BD}$ bisects ∠ABC

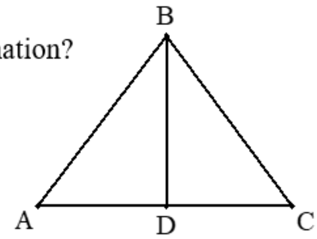

Figure 3: A Sample Task for the Drawing Conclusions Sub-Goal

Data Collection and Analysis

Pre-tests and post-tests were administered to 1,550 and 1,278 students, respectively, for three years of the study. The pre-test was a 20-item multiple choice assessment developed by Usiskin and used for Usiskin's and Senk's studies in the 1980s (e.g., Usiskin, 1982; Senk, 1985). The pre-test is called the Entering Geometry Test (EGT). The post-test, called the Cognitive Development and Achievement in Secondary School Geometry (CDASSG), was developed by Senk (1985). The CDASSG is a 6-item assessment. The first task is a fill-in proof. The second task provides a mathematical statement (e.g., The diagonals of a rectangle are congruent) and then requires only a Figure, a Given Statement, and a Prove Statement, but no proof. The last four tasks are full proofs. There are three forms of the CDASSG assessment that were designed to be approximately equivalent in difficulty.

Psychometric analyses of both instruments were carried out using classical true-score methods (Cronbach, 1951) and Item Response Theory (IRT; Lord, 1980). Descriptive item statistics were calculated, with percent correct used to quantify the difficulty of each item, while item-total correlations were used to confirm coherence of responses across items. Cronbach's Alpha was used to estimate the reliability of total test scores. Additionally, a 3-parameter logistic IRT model was estimated to produce item parameters (i.e., difficulty, discrimination, and guessing parameters), as well as item characteristic curves, a total information curve, and conditional standard errors of measurement for both assessments. For the CDASSG assessment, which included partial credit items, the IRT model was extended to a 3-parameter graded response model.

A total of 1,161 students completed both the pretest and posttest. Pearson's correlation between pre-test and post-test was calculated to confirm positive correlation between the two tests, and to serve as evidence of predictive validity. Approximately 1/6 of the sample was in the experimental group (n = 212) while the remaining students (n = 949) were in classrooms that used the standard curriculum. The students in the experimental group came from 4 of the 20 study teachers' classrooms (i.e., 4 teachers received the treatment and 16 did not).

Hierarchical Linear Modeling (HLM) was used to estimate impacts of the experimental curriculum. The HLM model is a random intercept analysis of covaraince (ANCOVA) model, with the following mathematical form.

$$Y_{ij} = \beta_0 + \beta_1(X_{ij}) + \beta_2(G_{ij}) + \beta_3(T_j) + \alpha_j + \varepsilon_{ij}$$

Where: Y_{ij} is the post-test score for student *i* under teacher j

β_0 is the model intercept

β_1 is the regression coefficient for the pretest score, X_{ij}

β_2 is the regression coefficient for an 8th Grade indicator (with G_{ij}=1 for students in 8th grade; G_{ij} = 0 for students in 9th grade or above)

β_3 is the regression coefficient for the treatment effect (with T_j=1 for experimental classes; T_j = 0 for non-experimental classes)

α_j is the random intercept for teacher j, distributed as $N(0, \tau)$

ε_{ij} is the residual term for student *i* under teacher j, distributed as $N(0, \sigma^2)$

Use of an ANCOVA-type model allows comparison of post-test scores under experimental and standard curriculum, after controlling for any preexisting differences in students' pretest scores on the Entering Geometry Test as well as differences associated with enrollment in an advanced mathematics class in 8th grade.

RESULTS

Classical psychometrics confirmed good reliability of both instruments, with Cronbach's Alpha of .82 for the Entering Geometry Test (EGT) and .86 on the CDASSG. The correlation between the pre-test (EGT) and post-test (CDASSG) is moderate to large (i.e., 0.67). Consequently, EGT scores can be used to predict CDASSG scores within ±20 Normal Curve Equivalent (NCE) points (i.e., approximately ±1 standard deviation). Figure 4 shows a scatterplot reflecting the relationship between pre-test and post-test scores.

Results from the IRT models also confirmed good reliability and validity of both instruments. Nearly all items had appropriate IRT difficulty estimates (e.g., between -2 and +2), good discrimination (e.g., above .50), and low guessing parameters (e.g., below .30). Total information and conditional standard errors of measurement suggest that both instruments have good precision for all students in the sample except those with unusually low (e.g., >2SD below the mean) or unusually high performance (e.g., >2SD above the mean).

The HLM models of treatment effect revealed a significant positive impact of the experimental curriculum. After controlling for grade level and pre-test (EGT) scores in a 2-level HLM with students nested within teachers, students in the treatment group scored 4.27 Normal Curve Equivalent points higher (Effect Size = +.20 standard deviations) on the post-test (CDASSG; p<.01). A second analysis was run because research suggests that teachers need time to learn how to teach from, and to develop trust in, new curriculum programs (Drake & Sherin, 2009). After restricting the HLM analyses to Year 2 (pre-implementation) and Year 4 (post-implementation) data, students in the treatment group scored 6.61 NCE points higher (ES = +.31 standard deviations) on the post-test (p<.001). Thus, the exclusion of the second year of data confirmed that impacts were larger after the initial implementation year. Taken together, both analyses demonstrate that the gains made by students were significantly larger with the GPS-focused lessons than the standard curriculum programs. Figure 5 illustrates the differences in post-test scores after controlling for pre-test scores.

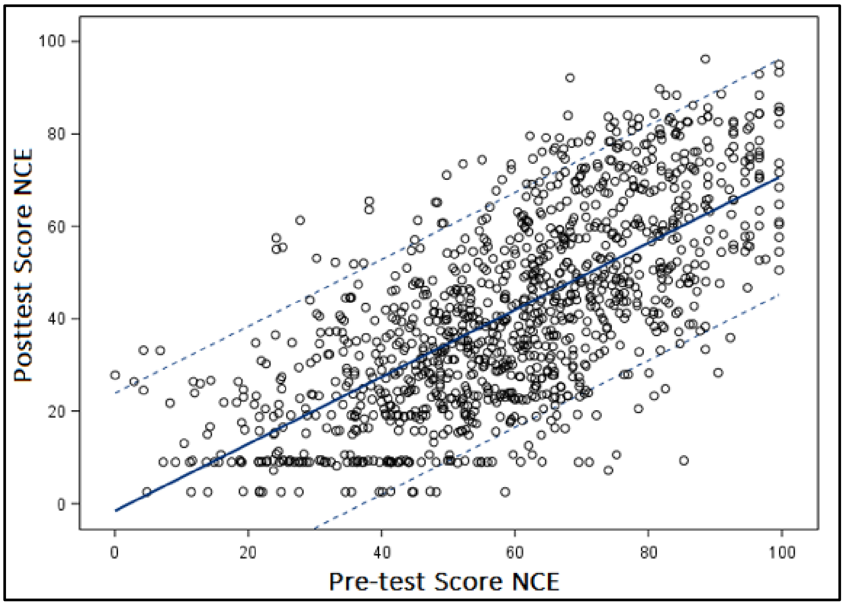

Figure 4: Scatterplot of pre-test and post-test scores (N = 1,161; r = .67)

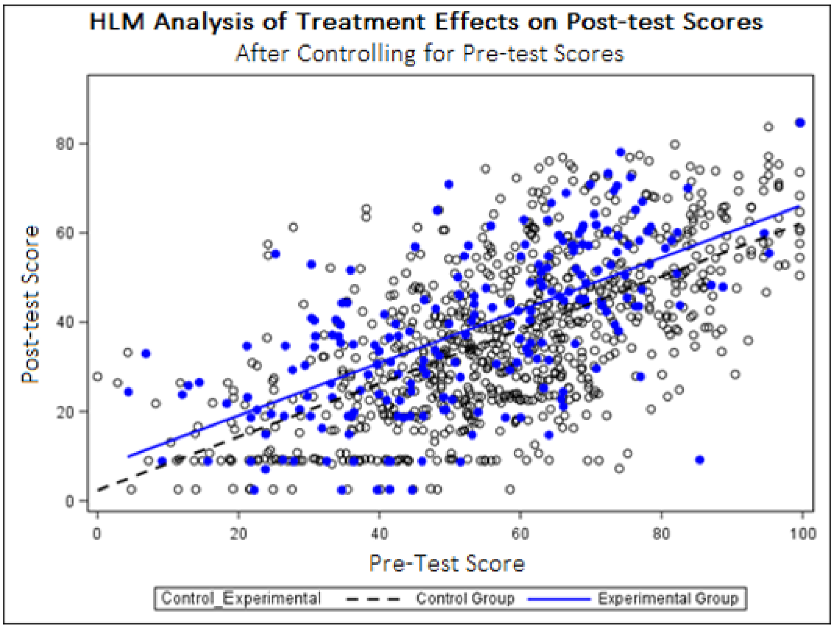

Figure 5: Scatterplot for Treatment Impact Analysis (N = 1,161; ES = +.20SD)

DISCUSSION AND CONCLUSION

We began this paper citing questions posed in Battista's (2007) review of the teaching and learning of geometry, such as: "What components of proof are difficult for students and why?" and "How can proof skills best be developed in students?" The results reported here provide a general proof-of-concept of an intervention for teaching proof in geometry. This intervention is based on the hypothesis that not only can pedagogical practices be decomposed, as put forth by Grossman et al. (2009), but so too can mathematical processes such as proving. While these results are promising, what is still unclear is which of the sub-goals and competencies included in the GPS-focused lessons are the true "high-leverage" practices, whereby "high-leverage" practices are defined as a "set of practices that have the greatest impact on student learning" (Hlas & Hlas, 2012, p. 78; O'Flaherty & Beal, 2018). The classroom teachers who participated in the study seemed to believe that the lessons focused on Developing Geometric Modalities were of particular importance. Other teachers who participated in the previous study (see Cirillo et al., 2017) found the Drawing Conclusions tasks to be especially important. More work is needed to better understand which of the sub-goals and competencies of the GPS are truly "high-leverage" and under what conditions they must be developed so that students can successfully engage in proof in geometry. Of course, it is also possible that some important competencies are still missing from the framework.

A limitation of this study is that we did not include item analyses, and our evidence is purely quantitative. Future work should explore achieved competencies as well as common student errors. A separate sub-study that made use of qualitative analyses, however, did explore competencies and behaviors observed when students solved two particular geometry proof tasks with smartpen technology. More specifically, Cirillo and Hummer (2021) found that students who were successful with the both proof tasks exhibited the following competencies: (a) they productively attended to the "Given" information; (b) they used the diagram as a resource; (c) they knew their warrants and explicitly identified them as postulates, axioms, definitions, or theorems; (d) they demonstrated that they were thinking in a logical manner, and (e) they attended to important details while working through their proofs (e.g., attending to common sub-arguments and the "Prove" statements). There are clear connections between these findings and the sub-goals and competencies in the GPS. For example, productively attending to the "Given" information is related to the Drawing Conclusions sub-goal, and using the diagram as a resource is related to the Working with Diagrams sub-goal. Also related, Cirillo and colleagues (in press) found that classes of students who did particularly well on the CDASSG assessment learned to prove theorems through a thoughtful routine that gave students opportunities to develop competencies from the Conjecturing, Understanding Theorems, and Understanding the Nature of Proof sub-goals. We conclude that a combination of qualitative and quantitative analyses will help us make further progress on Battista's important questions.

Acknowledgment

The authors would like to thank the teachers and their students for allowing us to conduct this research. They would also like to thank Jenifer Hummer and Amanda Seiwell for supporting the data collection and analysis. The research reported in this paper was supported with funding from the National Science Foundation (NSF; Award #1453493, PI: Michelle Cirillo). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

References

Battista, M. T. (2007). The development of geometric and spatial thinking. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 843-908). Charlotte, NC: Information Age Publishing.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cirillo, M. (2008). *On becoming a geometry teacher: A longitudinal case study of one teacher learning to teach proof*. PhD dissertation, Iowa State University.

Cirillo, M., Griffin, C., Seiwell, A., & Hummer, J. (in press). *"What do you believe is true?" A routine for proving theorems in secondary geometry*. Paper presented at the Proceedings of the 43rd Meeting of the North American Chapter of the International Groups for the Psychology of Mathematics Education Philadelphia, PA. Available at: https://www.pisc.udel.edu/references/

Cirillo, M., & Herbst, P. (2011). Moving toward more authentic proof practices in geometry. *The 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*. doi:https://doi.org/10.1007/s11858-021-01221-w

Cirillo, M., Murtha, Z., McCall, N., & Walters, S. (2017). Decomposing mathematical proof with secondary teachers. In L. West (Ed.), *Reflective and collaborative processes to improve mathematics teaching* (pp. 21-32). Reston, VA: NCTM.

Drake, C., & Sherin, M. G. (2009). Developing curriculum vision and trust: Changes in teachers' curriculum strategies. In J. T. Remillard, B. Herbel-Eisenmann, & G. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 321-337). New York: Routledge.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111*(9), 2055-2100.

Hlas, A. C., & Hlas, C. S. (2012). A review of high-leverage teaching practices: Making connections between mathematics and foreign languages. *Foreign Language Annals, 45*(s1), s76-s97.

Kobiela, M., & Lehrer, R. (2015). The codevelopment of mathematical concepts and the practice of defining. *Journal for Research in Mathematics Education, 46*(4), 423-454.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Flaherty, J., & Beal, E. M. (2018). Core competencies and high leverage practices of the beginning teacher: a synthesis of the literature. *Journal of Education for Teaching, 44*(4), 461-478. doi:10.1080/02607476.2018.1450826

Senk, S. L. (1985). How well do students write geometry proofs? *The Mathematics Teacher, 78*, 448-456.

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). Reston, VA: National Council of Teachers of Mathematics.

Smith, R. R. (1940). Three major difficulties in the learning of demonstrative geometry. *Mathematics Teacher, 33*, 99-134, 150-178.

Usiskin, Z. (1982). van Hiele levels and achievement in secondary school geometry. Chicago, IL: The University of Chicago. Retrieved from http://ucsmp.uchicago.edu/resources/van-hiele/