Eurasia Journal of Mathematics, Science & Technology Education, 2011, **7**(3), 181-192

Considerations Concerning Balacheff's 1988 Taxonomy of Mathematical Proofs

Thomas Varghese
Northern Alberta Institute of Technology, Edmonton, CANADA

Received 24 December 2009; accepted 23 April 2010

Current school curriculum documents require that justification and proof become a significant part of the mathematics classroom culture. In order to determine how well secondary level student teachers can complete a valid mathematical proof the researcher administered, the same mathematical task of Balacheff (1988) to a group of student teachers who were at the last semester of their teacher education program. The student teachers' written responses were then classified using Balacheff's Taxonomy of Proofs (BToMP). To assist in classifying student teachers' work, the researcher generated examples corresponding to Balacheff's taxonomy of proof. The purpose of this article is to confront the results of Balacheff and also to determine the various levels of proficiency with which the student teachers approached the task on the basis of BToMP. Along with the analysis of the results, the difficulties that the researcher encountered in categorizing student teachers' written work according to BToMP, for the same task he administered in his study is also discussed in this article This study raises questions concerning the applicability of BToMP, especially with advanced level students who have preconceived ideas about what would constitute a "preferred" approach to the proving task. It also suggests a need for further research into the thought processes and cognitive skills that are necessary, no matter what one's age, in solving mathematical proof tasks

Keywords: Balacheff's taxonomy of mathematical proofs, student teachers, mathematical proof

INTRODUCTION

Knuth (1999, 2002 a, 2002b) has noted that mathematical proof has played a peripheral role at best in North American secondary school mathematics education. He also observed that teachers introduce students to mathematical proof solely through the vehicle of Euclidean geometry. Several other researchers also noted the same phenomenon (Mariotti, 2000; Marrades & Gutierrez, 2000; Moore, 1994; Solomon, 2006; Sowder & Harel, 1998; Usiskin, 1987). Given this narrow application, the teaching and learning of mathematical proof does not appear to be that successful (Chazan, 1993; Coe & Ruthven, 1994; Healy & Hoyles, 2000; Hadas, Hershkowitz & Schwarz, 2000, Weber, 2001) in North America.

Curriculum documents published by National Council of Teachers of Mathematics (NCTM), a US based teacher association, stipulated that justification and proof are to become a significant part of mathematics classroom culture (NCTM, 1991 & 2000). This will be a challenging mandate. Jones (1997) notes that the teaching of mathematical proof places significant demands on the subject matter knowledge and the pedagogical knowledge of secondary mathematics teachers; and Knuth (2002 a, 2002b) insists that a teacher's conception of proof will influence both the role that mathematical proof comes to play within that teacher's classroom and the manner in which it is

Correspondence to: Thomas Varghese, Ph.D, Northern Alberta Institute of Technology, 11762-106 St. Edmonton, T5G 2R1, Alberta, CANADA
E-mail: thomasv@nait.ca

ISSN: 1305-8223

State of the literature

- Balacheff based his 1988 study on the mathematical practices of secondary school students as they encountered mathematical proof
- He took an experimental approach that allowed him to observe 28 children while performing a proof task
- The student responses were classified into four different types of proofs and argued that these categories represented four increasingly sophisticated level of thinking
- Balacheff (1988) reasoned that students' understandings of mathematical justification are likely to proceed from the inductive toward the deductive and toward greater generality

Contribution of this paper to the literature

- This study was conducted among student teachers who were in the last year of their Bachelor of Education program
- The responses were classified according to Balacheff's Taxonomy of Proofs
- The paper examines the difficulties encountered in the categorization of the responses
- It is also noted that despite the neatness of well defined categories, it is always not so easy to categorize the proofs under this taxonomy

taught. His study with secondary school mathematics teachers found that even practicing teachers have only minimal knowledge about the role and function of mathematical proof within the mathematics classroom. If teachers themselves lack the necessary mathematical understanding to feel secure in their instruction of a concept, it is quite natural that it will not receive the emphasis that curriculum developers would wish, nor the instructional time that curriculum documents stipulate.

This article presents the analysis of the replication of an existing research with secondary school student teachers conducted in a large Canadian University. Balacheff (1988) proof task was administered to a group of 17 secondary (math major) student teachers. The purpose of this study was to confront the results of Balacheff and also to determine the various levels of proficiency with which these student teachers approach the task on the basis of Balacheff's Taxonomy of Mathematical Proof (BToMP). Along with the analysis of my results, the difficulties that the researcher encountered in categorizing student teachers' written work according to BToMP, for the same task he administered in his study is also discussed in this article. The tasks was administered at a point when most of the student teachers were in the final semester of their teacher education program; in two weeks time all were off to complete their final teaching practicum.

BALACHEFF'S STUDY

Balacheff based his 1988 study on the mathematical practices of secondary school students as they encountered mathematical proof. He took an experimental approach that allowed him to observe twenty-eight thirteen and fourteen-year olds as they worked in pairs to generate proofs. The students were asked to *"provide a means of calculating the number of diagonals of a polygon when you know the number of vertices it has"* (p. 221). Balacheff facilitated interaction between the partners by providing only one pen for each pair. Students were allowed to work on the problem until they arrived at a solution; but both students had to agree that their answer did, in fact, provide a solution to the problem *before* they could claim to have finished. In this study, Balacheff focused on both the reasons that students gave for selecting the examples that they did and how they used those examples. He was keen to understand the processes involved in arriving at the product, but he understood that social interaction could either assist or hinder students in arriving at a solution to the proof (p. 222). After analyzing the results, Balacheff classified the student responses into four different types of proofs and argued that these categories represented four increasingly sophisticated levels of thinking.

Actually, it is in the "Proofs of Refutations" (Lakatos, 1976) that one can find the origin of the BToMP. It is also important to notice that BToMP is based on (i) the theory of didatical situations (which led to the experimental design), (ii) the Lakatos's preposition about the epistemology of proof (which establishes a link between proving and knowing) and (iii) Piaget genetic epistemology (which allows shaping possible precursor of mathematical proof).

Almost a decade later, Simon & Blume (1996) argued that Balacheff (1988)'s hierarchy of proofs was, in fact, an extension of van Dormolen's (1977) taxonomy of proof (as cited in Simon & Blume, 1996). According to Simon & Blume, van Dormolen had differentiated among proofs by establishing three distinct categories: proofs that (1) focus on a particular example, (2) use an example as a generic embodiment of a concept, and (3) use general and deductive argument. Balacheff identified four categories of proofs with his taxonomy: (1) naïve empiricism, (2) crucial experiment, (3) generic example and (4) thought experiment. Balacheff argued that each of these four levels of mathematical proof could be classified within one of two broad categories that he termed *pragmatic* justifications and *conceptual* justifications. He called all justifications *pragmatic* when they focused on the use of examples, actions or showings. He called justifications *conceptual* when they demonstrated abstract formulations of properties and relationships among properties.

The first three levels in Balacheff's proof scheme are all examples of pragmatic justifications. In the case of *naïve empiricism*, the first level in Balacheff's taxonomy of proofs, the student arrives at a conclusion concerning the validity of an assertion on the basis of only a small number of particular cases. Balacheff exemplifies this level in his description of the efforts of school students Pierre and Mathieu. Working together, these boys examined a square, a hexagon and then an octagon. They concluded that they could arrive at the number of diagonals by dividing the number of vertices by two. In this example of naive empiricism, the boys checked the statement to be proved against a few particular examples and, on this basis, made a universal assertion. With *crucial experiment*—the second level in Balacheff's taxonomy of proofs—the student deals with the question of generalization after generating a claim based on a few examples by examining a case that is not very particular. If the assertion holds in the considered case, the student will conclude that it is valid. Balacheff illustrated critical experiment by referring to the efforts of Nadine and Elisabeth. These girls chose a polygon of many sides (15) believing that the assertion they came up with could be proved in this instance, then the assertion would be universally true. In other words, at the level of crucial experiment, the student checks the statement by means of a carefully selected example. A defining characteristic of crucial experiment is the intentionality of the student. In other words, deliberate choices must be made (Knuth & Elliot, 1998) in the selection of an example.

Notably, both naïve empiricism and crucial experiment deal with actual actions or showings; the main difference between the two rests with the status of the specific example that is selected to validate the assertion—the example used in crucial experiment proof is often based on carefully selected extreme cases. I came to know through my work that one is more able to distinguish between crucial experiment and generic example while observing the student as s/he actually works through the task.

In the case of generic example—the third level in Balacheff's taxonomy of proofs—the proof rests upon the properties. Here, the example is a generalization of a class, not a specific example. Although the focus is once again a particular case, it is not used as a particular case, but as an example of a class of objects. The student selects such an example as representative of the class and performs operations/transformations on the example in order to arrive at a justification. Then, the student applies these operations and transformations to the whole class. Balacheff mentions Georges' exploration of the proposition $f(n) = n*s(n)$ (where $s(n)$ is the number of diagonals at each vertex) as an example corresponding to this category. However, it is quite interesting to note that Balacheff is not explicit in explaining his reasoning as to why this should represent a generic example (see Balacheff, 1988, pp. 224-225).

Only with the fourth and highest level of proof in Balacheff's taxonomy do students move from the practical—pragmatic justification—to the intellectual—conceptual justification. At the level of *thought experiment,* students are able to distance themselves from action and make logical deductions based only upon an awareness of the properties and the relationships characteristic of the situation. At this level, actions are internalized and dissociated from the specific examples considered. The justification is based on the use of and transformation of formalized symbolic expressions. Balacheff provides an example of thought experiment in his description of Olivier. This student asserted that *"In a polygon if you have x vertices there are automatically y diagonals from each point because in a boundary of the polygon there are two points which join it; in conclusion there are x-3 which are the diagonals"*. Olivier was able to express the properties of a polygon by observing one specific example. It is important to note that Balacheff categorized all assertions that de-contextualize themselves from the traces of formulation of their arguments, even if not necessarily fully correct, as thought experiments. In other words, it is the students' approach to the task of proving that he is categorizing not the validity of the outcome.

Knuth & Elliott (1998) explored BToMP further by providing examples to demonstrate each of the four levels of thinking. They used power chord theorem in their efforts to show how students thinking at any one of these levels might approach the task of proving the proof. In examining both Balacheff (1988)'s original study and the work of Knuth and Elliott, it has become clear to me that distinguishing between naive empiricism and crucial experiment is a difficult task, especially if one looks only at the end product of the student's engagement with the task. One could also observe that Knuth and Elliott did *not* provide a concrete instance of generic example. Hence, despite the neatness of the Balacheff model, in practical application one may have some difficulty both in distinguishing between naive empiricism and crucial experiment and in coming across instances of generic example.

Stylinadies (2007) provides an example of student argument, categorizing it as generic argument. Stylinadies phrased the students' argument as follows:

"If I take a number, say 200, and subtract it from itself, I get 0. Then if I add 10 to 0, I get 10. Because the same will hold for any number I choose to begin with, and because there is an infinite set of numbers I can choose from, there is an infinite number of answers I can choose from, there is an infinite number of answers for the problem "write number sentences that equal 10". (p. 312)

If the student had just written 200-200+10 = 10, and hence noted "there are infinite number sentences for 10", that could have been considered as a "naïve empiricism with a single example –and not a generic example. One could also argue that, as the study participants are primary school children, 200 is a "big number" to them, and hence "200" can be considered as an "extreme example", one that was "intentionally chosen". Then it could easily fit under the "crucial experiment" category. However, because of the explicit argument "*the same will hold for any number I choose to begin with, and because there is an infinite set of numbers I can choose from, there is an infinite number of answers I can choose from, there is an infinite number of answers for the problem",* (emphasis added) this argument could be clearly categorized as a generic example. This is the difficulty with most of the student's written work. Unless clearly indicated, a researcher can draw only limited conclusions based on the work alone.

The four levels in BToMP represent a hierarchy through which students are expected to progress as their notions of mathematical justification develop. Balacheff (1988) reasoned that students' understandings of mathematical justification are likely to proceed from the inductive toward the deductive and toward greater generality. Hence, those with increased mathematical maturity are most likely to be the students who generate deductive proofs. He also stressed that students will move back and forth between inductive and deductive reasoning depending on the task that they are completing. In other words, a student capable of thought experiment in one situation may regress to naïve empiricism in another.

The Study

Balacheff (1988)'s study centered on the *learning* of mathematical proof; the presented study began with an investigation of the learning of mathematical proof with the aim of ultimately enhancing the *teaching* of mathematical proof. The proficiency levels of student teachers who were asked to construct valid mathematical proof, were first examined; but because one's proficiency at a task is likely to influence the way in which one later instructs that task (Kunth, 2002a; Jones, 1997), the reseracher saw a critical relationship between proficiency and confidence: highly proficient learners of mathematical proof are likely to become more confident teachers of mathematical proof. It is commonly believed that a teacher's confidence in, and outlook on, mathematical proof will shape his/her teaching of these concepts (Fennema & Franke, 1992; Thompson, 1984).

The participants were informed ahead of time the objective of administering this task. A week after the information session, the task was administered. Of the 20 students present in the class, 3 decided not to participate in the study. The ethics committee of the University had approved the study. This study group consists of 17 Mathematics (major) students in the final semester of a teacher education program at a Canadian university. The teacher education in this particular university consists of either a 5 year combined program or a six year (after-degree) program. In the six-year program, students who had successfully completed a relevant first degree then applied to complete a 2-year BEd (Bachelor of Education) degree. For specialization in a secondary school subject, students must have had complete at least 12 three-credit courses in their subject of specialization. Hence, the participants of this study had completed a minimum of 12 university mathematics courses, including two courses in calculus as well as courses in geometry, linear algebra and abstract algebra.

ANTICIPATED PROOF CATEGORIES

In the following section, examples of how one might successfully complete the task *"provide a means of calculating the number of diagonals of a polygon when you know the number of vertices it has"* in accordance with BToMP is provided. The researcher noted that it is quite challenging to generate examples that would illustrate the various levels of thinking in this hierarchy of proofs. In particular, as noted before, it was difficult to generate illustrative examples of the approach categorized as generic example. The researcher understood from this study that unless the student uses some extreme examples (like a very large number or a polygon with a very many sides), or the student specifically mentions (aloud during the process or, perhaps, in writing) that the example is *intentionally* selected, it is quite difficult to differentiate between naïve empiricism and crucial experiment on the basis of written work alone. Balacheff's categories focus on the type of argument that the student presents rather than on whether or not the argument itself is correct; however, both factors are taken into consideration in the analysis of this article. First, the type of argument that the student has used is examined; and second, the success, or lack of success that the student has achieved in applying that argument is noted.

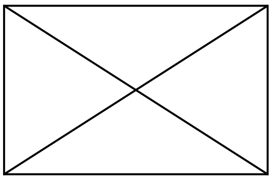
Exemplars

In the following section, exemplars corresponding to Balacheff's taxonomy of proof for the task: *"provide a means of calculating the number of diagonals of a polygon when you know the number of vertices it has"* is provided.

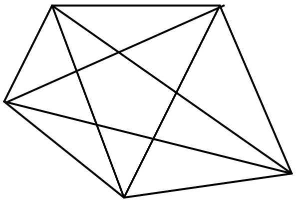
1. Approach: Naïve Empiricism

A rectangle has four vertices and two diagonals.

Vertices = v = 4
Diagonals = d = 2



A pentagon has five sides and five diagonals.



Vertices = v = 4
Diagonals = d = 2

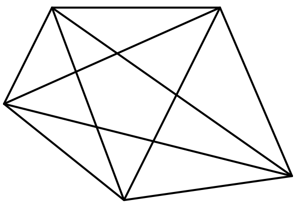
Hence, if "v" is even, the number of diagonals $d = v / 2$.
And if "v" is odd, the number of diagonals $d = v$.

2. Approach: Crucial Experiment

Balacheff (1988) distinguishes between crucial experiment and naïve empiricism on the basis of the student's selection of the example. Both Balacheff (1988, 1991) and Knuth & Elliot (1998) note that students at the level of crucial experiment intentionally select an extreme case, and if the proof works for that example, they will then conclude that their conjecture is correct and the proof proved. Thus, this task at the level of crucial experiment can be approached in the following manner.

I conjecture that the # of diagonals = # of vertices and will use the extreme case of the pentagon to verify my conjecture. I use the pentagon (as an extreme case) because it is the polygon with the greatest number of sides that I can still draw with relative ease.

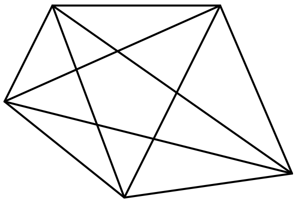
A pentagon has five sides and five diagonals



vertices = v = 5
diagonals = d = 5
Hence $d = v$

3. Approach: Generic Example

The pentagon has five sides and five diagonals.



Vertices = v = 5
Diagonals = d = 5

A pentagon has 5 sides (n = 5) and so 5 vertices (v = 5). From each vertex, one can draw only 2 diagonals because there are no diagonals from a vertex back to itself and there are no diagonals to the vertices on either side. Thus, there will be three fewer diagonals than the total number of sides (namely 2 at each vertex). Since there are 5 sides and 5 vertices, one can draw 5* 2 (= 10) diagonals in total. Diagonals have two ends; counting both ends of the same diagonal one would arrive at a total of 10. However, only one end needs to be counted. So, the number of diagonals will be 10 by 2, which equals the number of vertices.

This exemplar illustrates reasoning at the level of generic example because the calculations and answers are specific to the fact that one is considering a pentagon, although the same reasoning would apply whatever the number of sides involved. As noted above, Balacheff's (1988, 1991) example on generic reasoning does not include explicit rationales as to why the example constitutes the particular level of reasoning. When Knuth and Elliot (1998) expounded on Balacheff's proof levels using *power chord theorem* they also did not provide an explicit example and explained their rationale for their choice of illustration concerning generic example. Perhaps, one of the difficulties here lies in the specific problem task that Balacheff selected for his study (and which is the task for this study too): maybe it is simply too difficult to find a representative polygon that will work for all the different polygons of the same number of sides and all the different polygons. In other words, generic reasoning is easily verbalized, but when it is expressed on paper, it needs to be explained semantically in writing in order for one to assess the nature of the reasoning.

4. Approach: Thought Experiment

Arguing from the specific to the general distinguishes the generic example from the thought experiment. See the exemplar below:

Consider a polygon with "v" sides. If there are "v" sides, there are "v" vertices. Beginning with each vertex, one can draw (v - 3) (again, recall that there is no diagonal from a vertex back to itself and there are no diagonals to the vertices on either side). Thus, there will be three fewer diagonals than the total number of sides—that is, (v - 3) diagonals from each vertex. As there are "v" sides, there will be a total of $v(v - 3)$ diagonals. This approach, however, counts both ends of the diagonal. That means each diagonal is counted twice. Hence, to get the correct number of diagonals, divide the product by 2. Therefore, the formula for the number of diagonals is $d = v(v - 3) / 2$.

Analysis of the Task

As noted before, Balacheff (1988) did not take into consideration the correctness or incorrectness of the mathematical proof when placing it at a particular level of reasoning. However, student's work can be classified in one of three ways: as fully mathematically correct; as partially mathematically correct; and as mathematically incorrect. In this study, student efforts are categorized as correct only if the work is fully mathematically correct. In the case of this particular proof task (as noted above), the work of the two students who took the approach of naive empiricism can only be categorized as partially correct. (Knuth (1999), in his study, classed all proofs under the category of "naïve empiricism" as invalid.) In the analysis of this article, divide the thought experiment proofs into two categories a) verbal and b) symbolic. All efforts that are sophisticated: in other words, proofs that are enriched by the use of algebra or axioms or any other type of formalism and there is a minimal use of natural language (an example is provided later), are classified as thought experiment-symbolic. Since Balacheff's study group had limited exposure to mathematical proof, the examples of thought experiment in Balacheff's (1988) study are not very mathematical but verbal. In contrast, the present study group consisted of adults who could be considered mathematically sophisticated. Not surprisingly, then, they seem to have an image of proof form in mind. Hence, an additional category of "thought experiment-symbolic" is included.

From the table above, it can be noted that only 6 students out of 13 (4 provided correct formulas) were able to provide correct proof for the given task. Notably, other studies (Chazan, 1993, Coe & Ruthven, 1994, Healy & Hoyles, 2000; Weber 2001, Knuth, 2002a, 2002b) report that study participants found it difficult to complete correct proofs. When asked about their difficulties, most students pointed towards their experience with proof in secondary school. In secondary school mathematics, they experienced proof as compartmentalized within the domain of "Euclidean geometry". They also referred to a "discontinuity" between their proof experiences in secondary and post secondary schools. Generally in schools, students see proof as a "formal meaningless exercise", acquired by means of memory or the "received wisdom" bestowed upon them by their teachers. When they arrive at post secondary school, it is quite natural that they feel overwhelmed. What they had formerly little experience of was now presented as central to a mathematics education.

STUDENT WORK: PROOF CATEGORIES

In the following section, examples from student work that illustrate the different levels of mathematical reasoning as pertains to solving proof for Balacheff's (1988) task is presented. This task, then, involves the use of the definition of "diagonal". A majority of the students tried to solve the proof by using the definition. They examined various polygons in terms of their properties and gave an intuitive argument. This indicated that these student teachers had the conceptual tools needed to justify their argument. All of the student teachers worked with convex polygons as well, even though this was not a requirement stated in the question. This is similar to what Balacheff observed in his study.

Table 1. Observations from student work

Categories	*Correct*	*Partially Correct*	*Failed*	*Total*
Formula Only	4	0	0	4
Naïve Empiricism	0	2	0	2
Generic Example	2	1	0	3
Thought Experiment- verbal	3	0	3	6
Thought Experiment -symbolic	1	1	0	2
Total	10	2	3	17

The following sections discuss examples that illustrate Naïve Empiricism, Generic Example and Thought Experiment. It is to be noted that there were not any student work for this particular task that could be categorized as a "crucial experiment". The below given student names are all pseudonyms.

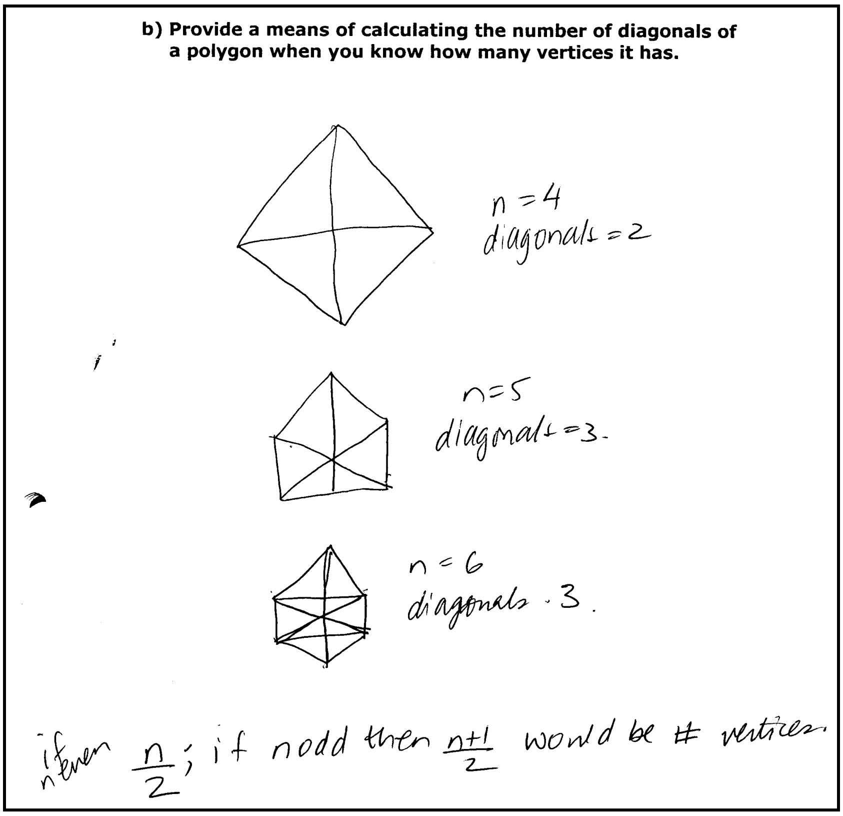


Figure 1. Tahira's solution

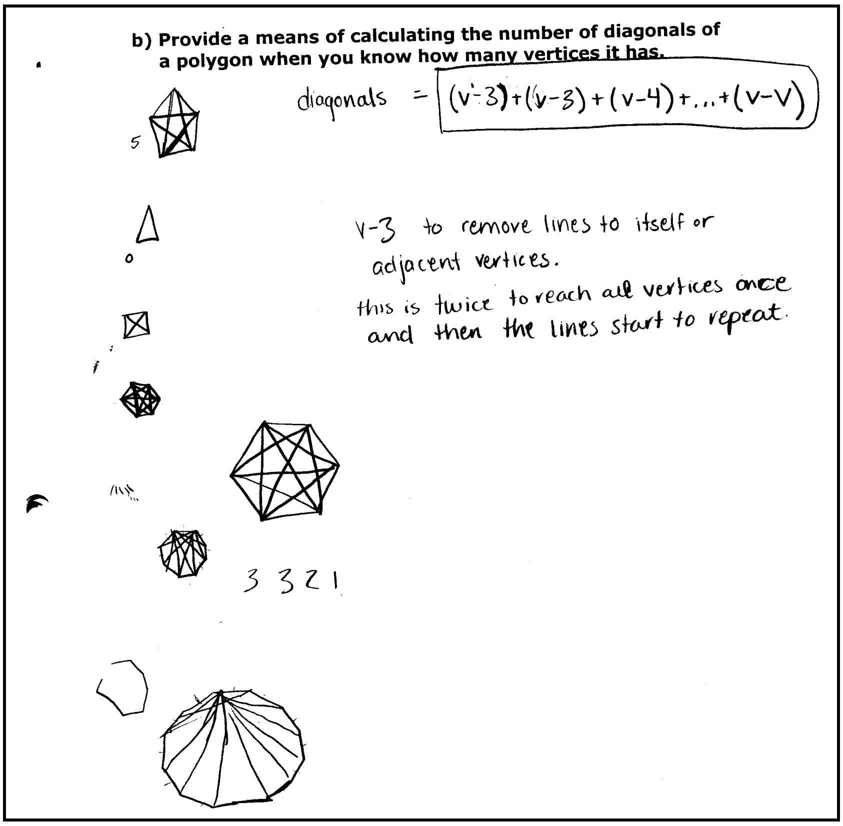


Figure 2. Grace's solution

Naïve Empiricism

Tahira explored(figure 1) three polygons—a square, a pentagon and a hexagon. She wrongly counted the number of diagonals for the pentagon and the hexagon; however, her incorrect counting fitted her conjecture very well since "If even $\frac{n}{2}$, if odd then $\frac{n+1}{2}$ would be # [of] vertices". There is another error in her conjecture. The question asks that she provide a means for calculating the number of diagonals, given the vertices. However, it seems that in her answer she provides a means for calculating the number of vertices. Her approach to proof suggests the first category in Balacheff's taxonomy as she explores a few polygons and comes up with a conjecture that she assumes to be true for all cases. This effort can be categorized as naïve empiricism. One could notice that Tahira's proof is similar to what Balacheff (1988) noted as Pierre and Mathieu's proof.

Generic Example

Assuming that she worked from top left to bottom right, Grace (figure 2) counted the number of diagonals for the pentagon and found it to be 5. She notes on the page, the number of diagonals next to the pentagon figure. Then, she seems to have moved to the triangle, noting that the number of diagonals is 0. Next, she likely drew a square and counted the number of diagonals, but did not bother to note the number. Then she moved to a hexagon and counted the number of diagonals correctly. She drew a heptagon and tried to count the number of diagonals, but left it undone. At that point, she seems to have tried an "extreme" polygon with 12 sides. However, she may have realized that it is not an easy job to count the number of diagonals of a 12-sided polygon. Whatever the reason, she left it undone after drawing only a few diagonals. It is apparent that she was trying to spot a pattern that would allow her to predict further results. Generating examples, looking for regularities in the data, making and articulating conjectures are the first steps towards generalization (Rowland, 2001).

It can be noted that Grace also draws another hexagon, bigger in size than the others, and with all the diagonals correct. This suggests that she returned to the hexagon and drew a larger diagram in order to make sense of the structure. This time she did not simply count the number of diagonals arriving at a sum of 9; instead, she wrote the structure as 3, 3, 2, 1. Since there are six sides in a hexagon there will be six vertices ($v = 6$); hence $(v – 3) = (6-3) = 3$ (there are two "3s" in the formula, "$(v - 3) + (v - 3)$"). The next number in her structure is "2", which agrees with her proposition "$v - 4$" = 6-4 and so on. "3 + 3 + 2 + 1" will yield the same result as "3 + 3 + 2 + 1 + 0"; hence, the formula is $(v - 3) + (v - 3) + (v - 4) + \ldots + (v - v)$ where "$v - v$" is 0. One could conjecture that she was attending to the hexagon when she wrote the formula. In other words, she uses the hexagon as a *generic example* in order to reach the general structure and the formula

$(v - 3) + (v - 3) + (v - 4) + \ldots + (v - v)$. It can be reasonably assumed that after she generated the formula for the number of diagonals, she wrote a partially complete general argument "v-3 to remove lines to itself [sic] or adjacent vertices. This is twice to reach all vertices once and then the lines start to repeat". Her argument is quite unclear to the researcher. One could suspect that she was trying to explain how she arrived at the formula.

Graces' work does not fit neatly into the category of generic example; furthermore, it contains traces of thought experiment This is the case with almost of the student work produced: if students experienced advanced level mathematics courses in which they became familiar with the form and elements of proof, then it makes sense that there will be traces of thought experiment in all of their proof forms.

Thought Experiment - Verbal

Chandelle (figure 3) tried 3 different polygons—square, pentagon and hexagon—in her efforts to make sense of the problem. She counted the vertices and the number of diagonals for each of these three polygons. Once she acquired a sense of the problem and the structure for generalization, she arrived at and justified a formula. She then verified her formula to determine whether or not she had arrived at the correct one. Since Chandelle uses the same examples that she had used earlier to explore the problem, I infer that she is now engaged in verification. Jahnke (2005) notes that some students will verify a statement, even after it has been proved, by means of examples (an observation, he claims, made also by Fischbein [1982]). I categorized this proof as a **correct verbal thought experiment** because Chandelle had developed a general explanation detached from the specifics of all her individual examples.

Thought Experiment - Symbolic

George (figure 4) made sense of the problem with the help of a triangle, square, pentagon, hexagon and heptagon. Even though he fails to specify his variables, it is evident that "v" stands for the vertices and "d" for the diagonals. The triangle does not have a diagonal;

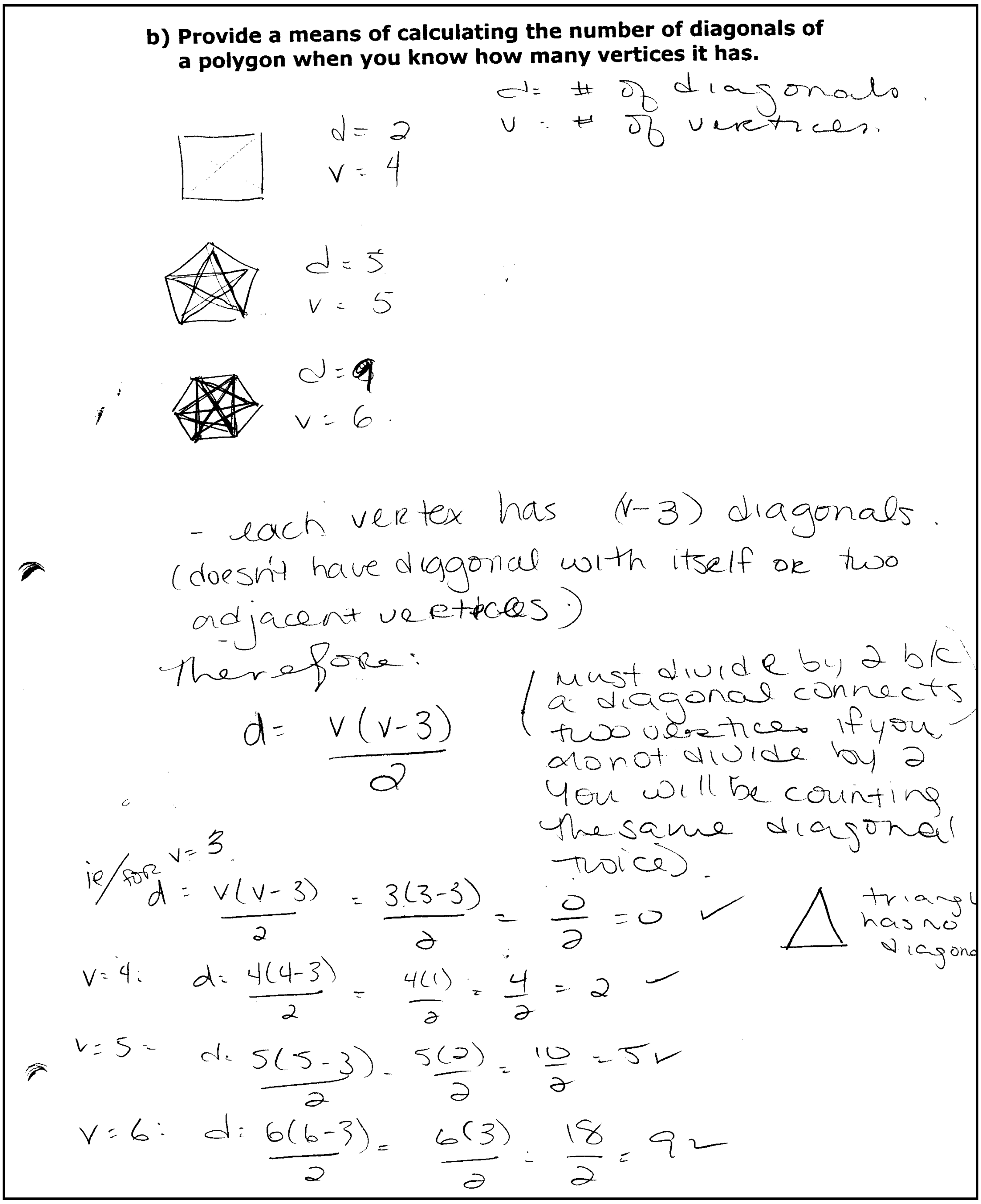


Figure 3. Chandelle's solution

hence, v = 3 and d = 0. Regarding the square, v = 4 and d = 2. The curved arrow pointing downwards and the +2 next to it indicate the difference in the number of diagonals between the triangle above and the square below. This pattern repeats with each new figure as the mathematical work proceeds down the page. When he arrives at the pentagon, he represents the number of vertices as v = 5 and the number of diagonals as d = 5. The difference in the number of diagonals between the square and the pentagon is 3; hence, he wrote +3. In the case of the hexagon, v = 6 and d = 9. The difference in the number of diagonals between the pentagon and hexagon is 4; hence, he wrote +4. In the case of the heptagon, v = 7 and d = 14. The difference in the number of diagonals between the hexagon and heptagon is 5; hence, he wrote +5. At the right hand side he writes "0", "0.5", "1", and "1.5" in an effort to establish a relation between vertices and diagonals. The pattern is provided below:

For triangle – $0v = d$
Square – $\frac{1}{2}v = d$
Pentagon – $v = d$
Hexagon – $3/2v = d$
Heptagon – $2v = d$

George did not go any further with the ratio relation between "vertices" and "diagonals". He could not arrive at a generalized formula with the ratio pattern, so he introduced a recursive relation $d_n = d_{n-1} + n - 2$. This relation is identical to that noted by Balacheff as f(n -1) + n - 2. It is interesting that George used "v" and "d" all the while, and then suddenly switched to "d" and "n". The way in which he spotted the pattern and the formula that he later developed from the pattern both represent sophisticated thinking. George put into play a number of different ideas in arriving at the general formula. At that point, it appears that he left that particular formula and, on the left side of the page, derived a formula for d_{n+1}. His derivation of $d_{n+1} = [(n + 1)^2 - 3(n + 1)] / 2$ is, in fact, correct and does yield the correct number of diagonals, for we substitute (n -1) for "n". It quite interesting that George made great efforts to derive dn+1, but then did not simplify this complicated expression. If simplified, the expression yields (n + 1)(n - 2)/ 2.

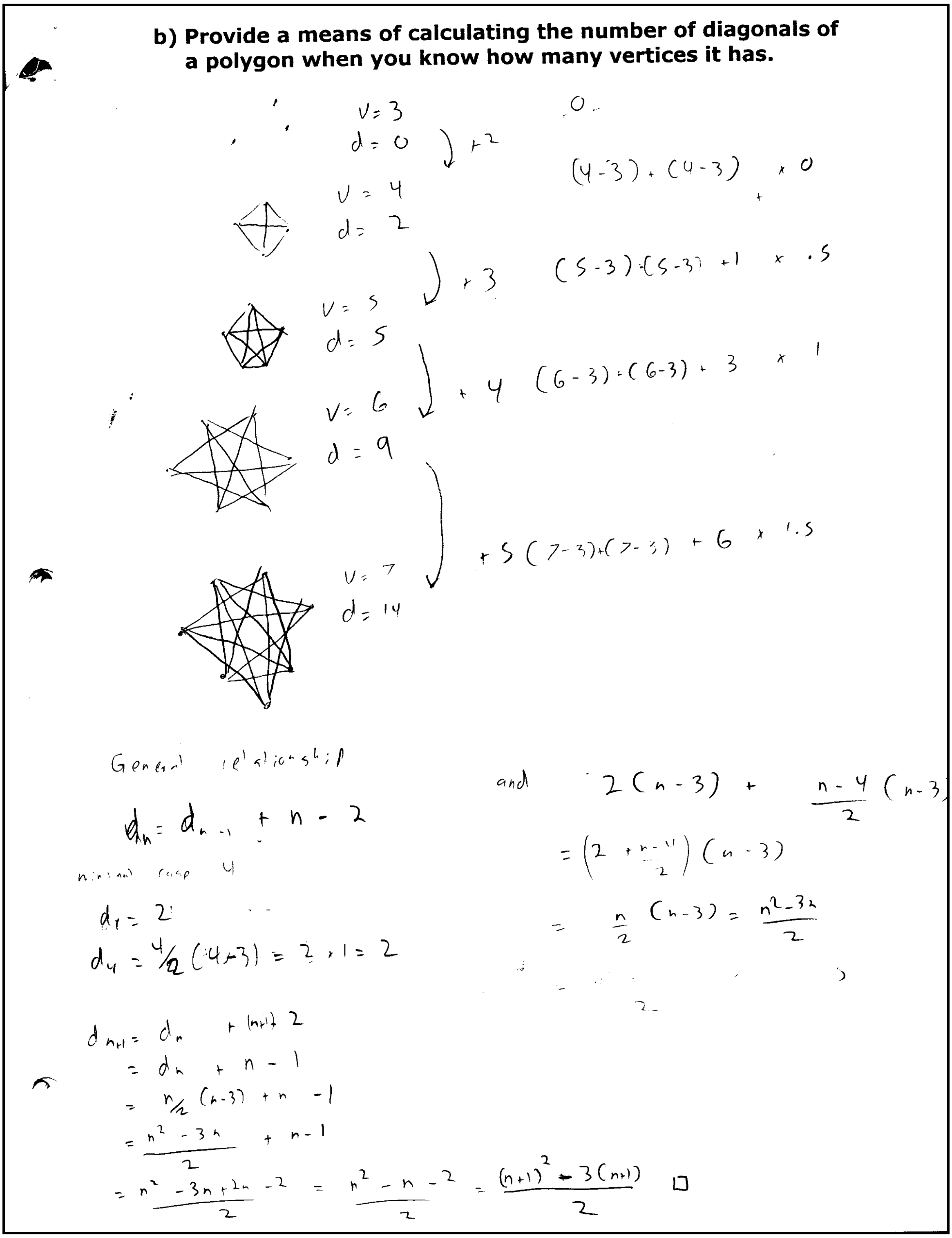


Figure 4. George's solution

On the right side of the recursive formula,

n/2 (n-3) = n2-3n/2

which is the correct formula for calculating the number of diagonals is also seen. He seems to have developed this as another expression for the number of diagonals. One can conjecture that 2(*n*-3) comes from what he initially observed (as noted in the upper right side)—that is, (as (4-3)+(4-3), (5-3) (5-3) and so on). The ratios 0, 0.5, 1, and 1.5 are transformed into and expressed as (*n*-4)/2 and the increment in the number of diagonals that occurs each time is (*n*-3). Thus, his formula is 2(*n*-3)+(*n*-4)/2*(*n*-3), which will yield $n/2(n\text{-}3)=n^2\text{-}3n/2$ when simplified. This is high level thought, indeed. George generates a mathematical proof based on induction and uses examples to demonstrate how the number of diagonals increases as the number of sides increase. His expressions demonstrate high-level mathematical thinking. Hence, this can be categorized as fully **correct mathematical proof.** George ended his proof with a hollow black square. This is a simple way of stating that the proof is complete. This symbol (or sometimes a dark black square (also called a tombstone) is usually used to end a proof when it is has been formally completed . In other words, this hollow black square is used instead of writing ***Q.E.D.*** which is an abbreviation of the Latin phrase *"quod erat demonstrandum"* (literally, "which was to be demonstrated"). Interestingly, only George and one other student ended their proofs in this formal way.

SUMMARY

Balacheff's (1988) study is one of the most often quoted in scholarly and professional publications dealing with mathematical proof. His work has influenced many researchers and his findings have long been a source of interest and debate. For all of these reasons, the researcher wished to conduct a study that would resonate with Balacheff's work. Hence, the researcher adopted his research design, carefully considered his findings, and devised a similar plan, though set it within a different context. As seemed to be the case with Balacheff, the researcher placed a great deal of importance on the concept of mathematical proof. Also like Balacheff, the goal of this study was to see how the participants engaged in the proving process. Balacheff gave his students ample time to complete the work and would only accept it when both students agreed that they had completed the task; the researcher also provided sufficient time for the completion of all tasks and permitted participants extra time if they needed it.

Based on the efforts of his teenage participants, Balacheff (1988) outlined a proof hierarchy reflecting four increasingly more sophisticated levels of thought and skill pertaining to mathematical proof. Moreover, he managed to place all of his students into one of these four levels with (apparently) relative ease. In this study, the researcher attempted to place each student participants into one of these four categories by BoTMP. The researcher worked with students who had completed a minimum of 12 university-level mathematics courses: their understandings of and experience with proof, likely, far exceeded that of Balacheff's thirteen- and fourteen- year-old participants. Since the student teachers in this study were already familiar with the expected forms for mathematical proof, they tended to try and make their work look 'mathematical.' It was noted that almost all of the proof work that these student teachers placed on the page reflected *traces* of thought experiment, the highest level in Balacheff's hierarchy of proof. But traces did not necessarily mean that the students had successfully and entirely reasoned through the problem at this, the most sophisticated level, within Balacheff's taxonomy. There was also evidence of lower levels of thought. Indeed, the data indicates that most of these young adults did not operate predominately nor successfully at the highest level in Balacheff's taxonomy of proofs. Hence, it was difficult to categorize their work and place their proofs into the *neat* categories afforded by Balacheff's well-defined taxonomy.

Balacheff's four levels of proof—naïve empiricism, the crucial experiment, the generic example, and the thought experiment—are developmental. Implicit in this hierarchy is the notion that students move from one level to the next, progressing to more mature and more sophisticated levels of thinking while embodying what has come before. The role of the teacher is to lead the students, by means of classroom discourse, towards higher and higher levels. This demands teachers to have excellent subject content knowledge and pedagogical knowledge.

It is also clear from my study that many students had considerable difficulty successfully completing proof tasks designed for the secondary school student. This finding corresponds with the findings of almost all of the studies that I examined (irrespective of the level of the participants) dealing with mathematical proof (Knuth, 1999, Coe & Ruthven, 1994, Healy & Hoyles, Martin & Harel, Chazan, 1993)

One particularly intriguing finding of this study is that even though one could neatly and clearly define Balacheff (1988)'s categories verbally, when it came to generating exemplars for these categories on paper, it is not that easy. Neither was it straightforward placing a particular proof format into one of these categories, at least the proofs completed by the participants in this study, students with significant mathematical proof experience.

Further research needs to be done on students' cognitive skills and reasoning processes as they work through mathematical proving tasks. When researchers in mathematics, as well as mathematics professors and professors of mathematics education, understand more clearly how students think through as they solve proofs, they will be better able to ensure that future elementary and secondary school mathematics teachers possess, not only the pedagogical knowledge of how to teach their subject, but the mathematical knowledge and *mathematical self-knowledge* that will enable them to do so with confidence and enthusiasm. The success of current reforms in mathematics education may depend upon it.

REFERENCES

Balacheff, N. (1988). *Aspects of proof in pupils' practice of school mathematics.* (In D.Pimm (Ed.), Mathematics, teachers and children (pp. 216-235). London: Hodder and Stoughton).

Balacheff, N. (1991). Treatment of refutations: Aspects of the complexity of a constructivist approach to mathematical learning. In E.von Glassersfeld (Ed.) *Radical Constructivism in Mathematics Education*, pp. 89-110. Dordrecht: Kluwer Academic Publishers.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics,* 24, 359-387.

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal.* 20, 41-53.

Fennema, E., & Franke, M. (1992). Teachers' knowledge and its impact. (In D. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning. (pp.147-164). New York: Macmillan.

Fishbein, E. (1982). Intuition and proof. *For the Learning of Mathematics,* 3(2), 9-24.

Hadas, N., Hershkowitz, R., & Schwarz, B. B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environments. *Educational Studies in Mathematics.* 44 (1-2), 127-150.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education,* 31(4), 396-428.

Jahnke, N. H. (2005). A genetic approach to proof. Retrieved January 10, 2007 from www.lettredelapreuve.it/ Newsletter/05Automne/ CERME4Jahnke.pdf

Jones, K. (1997). Student teachers' conceptions of mathematical proof. *Mathematical Education Review. 9. 23-32.*

Knuth, E. J. (1999). *The nature of secondary school mathematics teachers' conceptions of proof* . Unpublished doctoral dissertation. University of Colorado, Denver, Colorado.

Knuth, E. J. (2002a). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education.* 33, 379-405.

Knuth, E. J. (2002b). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education.* 5. 61-88.

Knuth, E. J., & Elliot, R. L. (1998). Characterizing students' understandings of mathematical proof. *Mathematics Teacher,* 91(8), 714-717.

Mariotti, M. A. (2000). Introduction to proof: The mediation of a dynamic software environment. *Educational Studies in Mathematics.* 44, 25-53.

Marrades, R., & Gutierrez, A. (2000). Proofs produced by secondary school studentslearning geometry in a dynamic computer environment. *Educational Studies in Mathematics.* 44, 87-125.

Martin, W. G. & Harel, G. (1980). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education.* 20, 41-51.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics.* 27, 249-266.

National Council of Teachers of Mathematics. (1991). Professional standards for teaching mathematics. (Reston: National Council of Teachers of Mathematics).

National Council of Teachers of Mathematics. (2000). Principles and standards of school mathematics. (Reston: National Council of Teachers of Mathematics).

Rowland, T. (2001). Generic Proofs: Setting a good example. *Mathematics Teaching,* 177. 40-43.

Simon, M., & Blume, G. (1996). Justification in the mathematics classroom: A study of prospective elementary teachers. *Journal of Mathematical Behavior.* 15(1), 3-31.

Solomon, Y. (2006). Deficit or difference? The role of students' epistemologies of mathematics in their interactions with proof. *Educational Studies in Mathematics.* 61, 373-393.

Sowder, L., & Harel, G. (1998). Types of students justifications. *The Mathematics Teacher,* 91, 670-675.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education.* 38 (3), 289-321.

Thompson, A G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics.* 15, 105-127.

Usiskin, Z. (1987). Resolving the continuing dilemmas in school geometry. In M. M. Lindquist & A. P. Shulte (Eds.), *Learning and teaching geometry, K-12 (pp. 17-31).* Reston, VA: National Council of Teachers of Mathematics.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics.* 48, 101-119.