INVESTIGATIONS IN MATHEMATICS LEARNING
©The Research Council on Mathematics Learning
Fall Edition 2013, Volume 6, Number 1

MIDDLE SCHOOL CHILDREN'S MATHEMATICAL REASONING AND PROVING SCHEMES

Yating Liu
Old Dominion University

Azita Manouchehri
The Ohio State University

Abstract

In this work we explored proof schemes used by 41 middle school students when confronted with four mathematical propositions that demanded verification of accuracy of statements. The students' perception of mathematically complete vs. convincing arguments in different mathematics branches was also elicited. Lastly, we considered whether the students recognized and identified advantages associated with using justification models different from their own in order to offer a theoretical account for how individuals' proof scheme choice might be impacted by such an exposure.

In mathematics, truths are established via proofs (Balaguer, 2008; Brown, 2008; Harel & Sowder, 2007; Jaffe & Quinn, 1993; Krantz, 2007; Lakatos, 1976; Schoenfeld, 1994). Indeed, it is this particular emphasis on proofs that sets mathematics apart from all other scientific or analytical disciplines (Fawcett, 1995/1938). Krantz (2007) positioned this point well:

> The unique feature that sets mathematics apart from other sciences, from philosophy, and indeed from all other forms of intellectual discourse, is the use of rigorous proof. It is the proof concept that makes the subject cohere, that gives it its timelessness, and that enables it to travel well. (p. 1)

Recognizing the vital role of proofs in the discipline and in school mathematics (Harel & Sowder, 2007), both the Principles and Standards for School Mathematics (NCTM, 2000) and the Common Core State Standards

(Council of Chief State School Officers, 2010) place tremendous emphasis on the need to assist school children in developing their proving skills (Boero, 2007; Hanna & de Villiers, 2012; NCTM, 2000; Stylianides, 2007; Stylianou, Blanton, & Knuth, 2009). Historically (and currently), in the US, a course on Euclidean geometry has served as the main venue for the development of students' skills in deductive reasoning with the expectation that such skills would automatically transfer to other mathematical and nonmathematical areas (González & Herbst, 2006; Herbst & Brach, 2006). This goal, however, remains unfilled. It is recognized that this failure might be due to the school treatment of topics in curriculum and instruction. There is evidence that in many mathematics classrooms proofs and the proving process are taught procedurally instead of as a conceptual tool for reasoning (Herbst & Brach, 2006; Reid, 2011). As a consequence, students tend to view proof as a special "form" of producing written work (e.g two-column proof) instead of a viable vehicle for production of reliable explanations, or even means for understanding (Chazan, 1993; González & Herbst, 2006; Schoenfeld, 1988). Existing literature unanimously suggests that when evaluating arguments students seem naturally and insistently to prefer empirical reasoning over deductive arguments (González & Herbst, 2006; Healy & Hoyles, 2000; Tall et al., 2012). Additionally, there is evidence that an understanding of the role of mathematical proofs in establishing validity of arguments remains underdeveloped at all grade levels (Chazan, 1993; Chazan & Lueke, 2009; Harel & Sowder, 1998; Heinze & Reiss, 2009; Kuchemann & Hoyles, 2009; Mason, 2009; Waring, 2000; Weber, 2001; Schoenfeld, 1988). Careful epistemological analysis, through research, on the factors that influence students' proving schemes and ways in which individuals' thinking and reasoning might be evolved, is needed so as to facilitate the development of a *Comprehensive Perspective* (Harel & Sowder, 2007, p. 1) on learning and teaching proofs (Marrades & Guitierrez, 2000). Our research was designed to address this need.

Grounding our work in Harel & Sowder's (1998) taxonomy of proof schemes and Balacheff's (1988) notion of epistemological beliefs as windows to understanding individuals' mathematical actions, in this work we examined three issues. First, we elicited the proof schemes used by students when verifying validity of statements regarding mathematical relationships from four different content areas. Second, we attempted to assess whether students' preference for alternative arguments they were offered differed from those they themselves had used. Third, we studied students' notions of *mathematical completeness* and *convincingness* according to the relationships they may have identified. Data collection and analysis was informed by four research questions:

1. What justification schemes do middle school students use when solving problem contexts that require proving?

2. Do middle school students' justification schemes differ according to the mathematical content of the task?
3. What type of justification scheme do middle school students prefer when offered alternative types of validation approach to tasks?
4. What is the relationship, from the middle school students' perspective, between *mathematically complete* and *mathematically convincing* arguments, and how may this perspective impact their preference of proof schemes?

Proofs and Proof Schemes

Proofs serve multiple purposes in mathematics. There is consensus among scholars of mathematics education and mathematicians that verification, justification, illumination, and systemization are key functions of proofs in mathematics (Bell, 1976; Harel & Sowder, 2007; Krantz, 2007). Verification and justification process concerns establishing the truth of propositions. Illumination implies that the proof itself conveys insight into why a proposition is true. The third function, systemization, is an attempt to organize the results into a deductive axiomatic system. More recently, de Villiers (2003) noted discovering new theorems, communicating and/or transmitting knowledge, and providing intellectual challenge to the author of the proof as additional objectives of proofs. Each of these functions is recognized to be crucial to school mathematics (Ball & Bass, 2003; Harel & Sowder, 2007).

Despite the critical role of proofs in the discipline, there is overwhelming evidence of students' difficulties with producing them. Marrades and Guitierrez (2000) noted two specific categories of research on proofs: "descriptions of the students' work when solving proof problems and descriptions of students' beliefs when deciding whether they are convinced by an argument about the truth of a statement" (p. 89). Indeed, the study of learners' proof schemes has a long history and is currently a main stream in didactics of mathematics. For instance, Bell (1976) identified empirical and deductive as two major modes of justification that students used when working on problems that demanded proving. According to his description, empirical justification relies on the use of examples whereas deductive justification relies on deduction to connect data with conclusions. Bell further catalogued students' answers according to the different degrees of completeness of checking the statement in the whole (finite).

Balacheff (1988) coined pragmatic and conceptual as two modes of justification prominently used by the students he studied. Pragmatic justifications are based on the use of examples (or an actions), and conceptual justifications are based on abstract formulations of properties and of relationships among properties. He further identified three types of pragmatic justifications to include: naive empiricism, in which a statement to be proved is checked using a few (somewhat randomly chosen) examples; crucial experiment, in which a statement is checked in a carefully selected example; generic example, in which the justification is based on operations or transformations on an example which is selected as a characteristic representative of a class. The category of conceptual justifications incudes thought experiment, in which actions are internalized and dissociated from the specific examples considered, and symbolic calculations from the statement, in which there is no experiment and the justification is based on the use and the transformation of formalized symbolic expressions (See Figure 1).

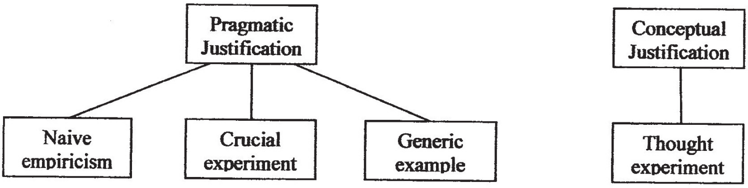


Figure 1. Proof Schemes of the Arguments in Each Problem

Balacheff (1988) concluded that while students might experience difficulty with producing proofs, they do show awareness of the necessity to prove using logical reasoning. Extending the research of Bell (1976) and Balacheff (1991), and drawing from their own empirical data, Harel & Sowder (1998) proposed their taxonomy of proof schemes to consist of three main categories, i.e. external, and analytical, each of which encompasses several subcategories (see Figure 2).

In particular, external conviction proof schemes include instances where students determine the validity of an argument by referring to external sources, such as the appearance of the argument instead of its content (e.g. they tend to judge upon the kind of symbols used in the argument instead of the embedded concepts and connection of those symbols), or words in a textbook or told by a teacher. Empirical proof schemes, inductive or perceptual, include instances when a student relies on examples or mental images to verify the validity of an argument; the prior draws heavily on examination of cases for convincing oneself, while the latter is grounded in more intuitively coordinated mental procedures without realizing the impact of specific transformations. Lastly, analytical proof schemes rely on either transformational structures (operations on objects) or axiomatic modes of reasoning which include resting upon defined and undefined terms, postulates or previously proven conjectures. Harel & Sowder also posited that students tend to maintain various understandings about what they are expected to do when asked to offer a proof, and these understandings might

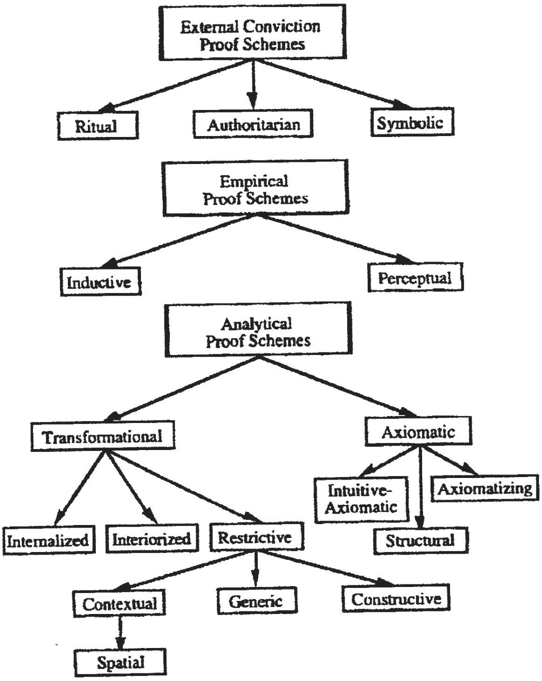


Figure 2. Proof Schemes and sub schemes (Sowder & Harel, 1998)

be strongly impacted by personal preference or "orientations" (Schoenfeld, 2011).

There is a consensus that in order for students to move towards relying on analytical schemes when attempting to validate mathematical statements, they need to recognize deductive arguments as convincing statements instead of merely mechanical procedures that need to be followed (Chazan, 1993; Tall et al., 2012). Despite this shared understanding, research has consistently shown that misconceptions about proofs impede students' capacity to recognize this critical feature (Harel & Sowder, 1998; Healy & Hoyles, 2000; Herbst & Brach, 2006). Recio and Godino (2001) coined the construct of *Personal Explanatory Argumentation Schemes* (PEAS) (pg. 97) to explain the nature of individuals' reasoning. They characterized PEAS as intuitive arguments that individuals use to explain relationships as opposed to validating them. Highlighting the experimental, subjective nature of this mode of reasoning, they offered that PEAS are deeply rooted in mathematical proofs as individuals first establish conjectures through personal schemes and intuitive procedures and then validate them using deductive proofs. As such, they cautioned against assuming these personal schemes as unimportant or primitive:

> Informal proof schemes should not be considered as simply incorrect,

> mistaken or deficient, but rather as facets of mathematical reasoning necessary to achieve and master mathematical argumentative practices. The analytical arguments, which are characteristic of mathematical proofs, are not the sole argumentation practices used by mathematicians to convince themselves about the truth of their conjectures. These reasoning procedures are often unfruitful, or even an obstacle, in the creative/discovery stages of problem solving processes, in which it is allowed and even necessary to implement substantial ways of argumentation in particular, empirical induction and analogy. (p. 93)

The authors concluded that it is necessary to link the different meanings of proof, at different teaching levels, so as to help students develop not only the skills but also the rationality required in applying proofs appropriately. Such a development demands that both the knowledge of proofs and the discriminative capacity to use them be nurtured in curriculum and instruction, progressively and over time.

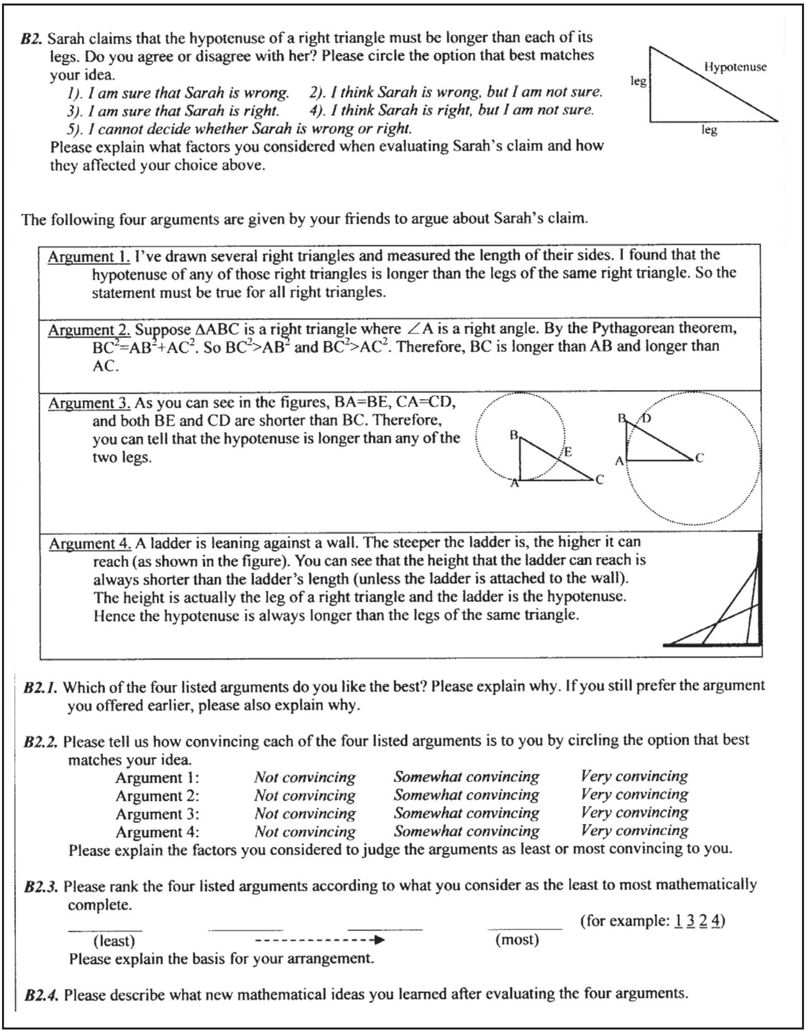
Success in designing instructional programs that enhance knowledge and capacity depends largely on the construction of a model of proving process that is grounded in a deep understanding of what different student populations at different stages of their mathematical development, might consider as proofs (Stylianides, 2007). The existing models have been built around thinking of upper secondary and college students' work, grounded primarily in proof-based courses such as geometry, linear algebra, discrete mathematics and real analysis (Harel & Sowder, 2007).

Research studies that explain proof schemes used by middle school children and/or their proving processes of individuals in more than one subject area are rare. Moreover, the existing models have been developed using a homogeneous sample from groups enrolled in the same course. As such, the students' performance could be attributed to exposure to a particular type of instruction or curriculum or common social practices among them. Our research was conceived to advance the field (Dreyfus, 2006) by including three novel contexts in the study. First, we investigated proof schemes of a group of eighth grade students. This particular population is critical since the grade band serves as the bridge between informal and formal levels of mathematical work of elementary and secondary school. Second, the participants were drawn from 19 different middle schools across the state of Ohio, suggesting variety in both content and heuristics they may have been experiencing at the time of data collection. Third, we examined the students' proving processes simultaneously in four different content areas as a means to inspect the potential relationship between the content of a problem and reasoning scheme that may have been elicited by it. These are described more fully in the following sections.

Method

Date Collection Instrument

In collecting the needed data for the study we developed a *Survey of Reasoning* (Liu & Manouchehri, 2012), which examined students' skills in producing mathematical proofs, their ability to judge validity of alternative external justification methods, and their willingness to adopt particular models of justification when forming mathematical arguments. We designed the survey items so as to meet the contents of the theoretical indi-



B2. Sarah claims that the hypotenuse of a right triangle must be longer than each of its legs. Do you agree or disagree with her? Please circle the option that best matches your idea.

1). I am sure that Sarah is wrong. *2). I think Sarah is wrong, but I am not sure.*
3). I am sure that Sarah is right. *4). I think Sarah is right, but I am not sure.*
5). I cannot decide whether Sarah is wrong or right.

Please explain what factors you considered when evaluating Sarah's claim and how they affected your choice above.

The following four arguments are given by your friends to argue about Sarah's claim.

Argument 1. I've drawn several right triangles and measured the length of their sides. I found that the hypotenuse of any of those right triangles is longer than the legs of the same right triangle. So the statement must be true for all right triangles.

Argument 2. Suppose ΔABC is a right triangle where $\angle A$ is a right angle. By the Pythagorean theorem, $BC^2=AB^2+AC^2$. So $BC^2>AB^2$ and $BC^2>AC^2$. Therefore, BC is longer than AB and longer than AC.

Argument 3. As you can see in the figures, BA=BE, CA=CD, and both BE and CD are shorter than BC. Therefore, you can tell that the hypotenuse is longer than any of the two legs.

Argument 4. A ladder is leaning against a wall. The steeper the ladder is, the higher it can reach (as shown in the figure). You can see that the height that the ladder can reach is always shorter than the ladder's length (unless the ladder is attached to the wall). The height is actually the leg of a right triangle and the ladder is the hypotenuse. Hence the hypotenuse is always longer than the legs of the same triangle.

B2.1. Which of the four listed arguments do you like the best? Please explain why. If you still prefer the argument you offered earlier, please also explain why.

B2.2. Please tell us how convincing each of the four listed arguments is to you by circling the option that best matches your idea.

Argument 1:	*Not convincing*	*Somewhat convincing*	*Very convincing*
Argument 2:	*Not convincing*	*Somewhat convincing*	*Very convincing*
Argument 3:	*Not convincing*	*Somewhat convincing*	*Very convincing*
Argument 4:	*Not convincing*	*Somewhat convincing*	*Very convincing*

Please explain the factors you considered to judge the arguments as least or most convincing to you.

B2.3. Please rank the four listed arguments according to what you consider as the least to most mathematically complete.

_______ _______ _______ _______ (for example: 1 3 2 4)
(least) ------------► (most)

Please explain the basis for your arrangement.

B2.4. Please describe what new mathematical ideas you learned after evaluating the four arguments.

Figure 3. A sample problem from the *Survey of Reasonng*

cators identified by Harel and Sowder's (1998) taxonomy of proof schemes, as described below. Due to the fact that a measurement tool was not available, when we commenced our project we also considered, in our research, whether the tool we developed and used was suitable for eliciting middle school students' thinking in pertinent areas.

Content of the Survey

The *Survey of Reasoning* consisted of four mathematics problems from four different branches of mathematics (i.e. number theory, geometry, probability, and algebra). Each problem consisted of several parts (See Figure 3 for an example).

First, the participants were presented with a proposition and asked to determine whether they agreed with and were sure about the accuracy and completeness of the statement. The options provided for them included: *I am sure the statement is wrong; I think the statement is wrong but I am not sure; I am sure the statement is right; I think the statement is right but I am not sure; I can't decide whether the statement is right or wrong.* They were also asked to offer an explanation for their choice and factors they considered when evaluating the statement. In the second part, four arguments, each embodying a different proof scheme supporting or refuting the same statement, were offered. The participants were asked to compare their own argument to those given, and to decide whether they preferred any of the optional statements over their own method. Lastly, they reported if they considered each of the optional arguments as *convincing* and/or *mathematically complete.* We deliberately chose the terms *convincing* and *mathematically complete* to evaluate students' "two conceptions of proof" (Healy & Hoyles, 2000), assuming that when judging the *convincingness* of an argument the students might tend to rely on subjective perceptions, whereas when judging the *mathematical completeness* they might refer to an understanding of existing mathematical conventions. Table 1 presents a blueprint of the types of proof schemes we utilized in the arguments used in each of the problems.

Tablel 1.
Proof Schemes of the Arguments in Each Problem

	Number Theory	Geometry	Probability	Algebra
Argument 1	Inductive	Inductive	Axiomatic	Transformational
Argument 2	Perceptual	Axiomatic	Inductive	Inductive
Argument 3	Axiomatic	Transformational	Inductive	Inductive
Argument 4	Perceptual	Transformational	Axiomatic	Axiomatic

We considered two major issues in the development process. First, we agreed that the proof schemes in Harel & Sowder's model (1998) are not mutually exclusive and a proof may simultaneously engage more than one scheme. In cases where the classification was not conclusive we categorized an argument based on our judgment of the proof scheme that was most crucial for development of a convincing argument. For instance, consider Argument 4 in the geometry problem (see Figure 3). On the one hand, perceptual reasoning was involved in the argument since a connection was made between the ladder scenario and the image of right triangles. On the other hand, transformation reasoning was also involved when considering the movement of the ladder (hypotenuse). One could recognize the generality of the argument by conceptualizing such a movement. Second, Harel & Sowder (1998) suggested a major difference between perceptual and transformational reasoning to include whether the problem solver is able to initiate the transformation in the argument. Regarding this identifier, we considered whether the movement of ladders was a common scene in real life and the problem solver simply referred to the scene to support his/her argument in the triangle problem, or the problem solver intentionally used the movement of the ladder to represent different shapes of right triangles. In this particular case, we accepted the latter and classified the argument as having a transformational proof scheme. We agreed that regardless of how carefully the proof scheme categories might be defined legitimate debates surrounding their borderlines may continue to persist. As a result, findings of the student should be considered with respect to these flexibilities.

Participants

The sample consisted of 41 eighth grade students from 19 different urban middle schools across the state of Ohio. The students had agreed to participate in a mathematics enrichment program offered through the Office of Minority Affairs, Young Scholars Program at the Ohio State University as an ongoing effort to recruit and prepare students from urban communities in STEM related initiative. The enrichment program was developed and implemented by the authors. At the time of data collection, approximately 40% of the participants were enrolled in a course in Geometry. The remaining students were enrolled in either an Algebra I course or in a course titled Pre-Algebra. The academic standing of the participants represented the diverse middle school population with one notable difference. The children enrolled in the enrichment program were from disadvantaged communities and considered first generation college attendees in case they ever chose to attend college. Their participation in the program was voluntary and no monetary compensation was offered to them. As such they served as a representative of children enrolled in eighth grade statewide.

The *Survey of Reasoning* was administered to all participants on the first day of the enrichment program and used as a pre-assessment tool for determining the nature of their mathematical reasoning skills. The participants were given 90 minutes to complete the Survey and allowed to use calculators if they were inclined to do so. It was reinforced, at the time of administration, that there was no right or wrong answer to any of the questions they were asked to answer and that the most important thing, from our standpoint, was for them to explain their ideas.

Data Analysis Process

Data analysis followed a three-stage process. First, using Harel and Sowder's (1998) taxonomy we coded the participants' responses to the first part of each of the survey items. We then identified both the type and the frequency of occurrence of a particular scheme that the participants had used for each of the problems and then across the four problems. Our analysis focused on categorizing participants' responses according to problem type and the contexts used on the survey. Results of the first stage of analysis were used to answer research questions 1 and 2.

At the second level of analysis, in answering the third research question we divided all 16 alternative arguments we had used on the *Survey of Reasoning* into two groups based on their underlying schemes. Each of the empirical and the analytical groups consisted of eight arguments. The proof schemes of students' preferred arguments were quantified in order to depict the potential instructional impact of the given arguments, in particular to identify whether the preferred arguments involved more analytical reasoning than those created by the participants. At the third level of analysis we studied whether the participants' understanding of *convincing* arguments and *mathematically complete* arguments were consistent with each other. In doing so, we assigned a *students' judgment of convincingness* (SJC) score of 1 to 3 to each participant's response to the corresponding survey questions (e.g., see Question B2.2, Figure 3), where 1, 2, and 3 indicates *not convincing, somewhat convincing,* and *very convincing,* respectively. Similarly, we assigned a *students' judgment of mathematical completeness* (SJMC) score of 1 to 4 to each participant's response to the corresponding questions in the survey (e.g., Question B2.3, Figure 3), where a lower score indicates the argument is less mathematically complete as judged by the participant. Level of association between the two scores was examined to answer research question 4. Findings were further supported by students' explanations to their options (see Figure 4 for a graphical illustration of the phases in data analysis).

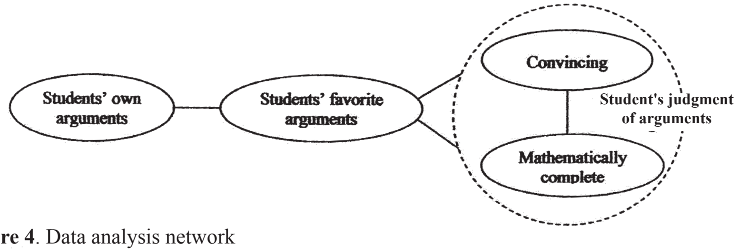


Figure 4. Data analysis network

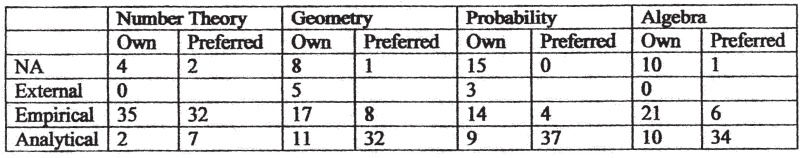
Results

In this study our goal as to document the validation processes used by a group of eighth grade students when determining the accuracy of mathematical propositions from four different content areas; to explore whether the reasoning schemes that the students preferred, based on examples provided for them, differed from arguments they themselves had used; and to trace the individuals' understanding of the roles and functions of mathematical proofs. Findings regarding each of these goals are discussed below.

Students' Modes of Reasoning

Table 2 offers a summary of the proof schemes students used and preferred when working on each of the four problems (Note: "NA" indicated that either students did not show any work that might contribute to solving the problem or the students didn't offer information that was adequate for the analysis). The data suggested that even in the context of the geometry problem where students were more likely to use analytical explanations than in the other problems, only 11 out of 41 students supported their conclusion analytically. This result implied that a majority of students relied heavily on reasoning with empirical and external reasoning schemes.

Table 2.
Students' Own Proof Schemes vs. Preferred Proof Schemes



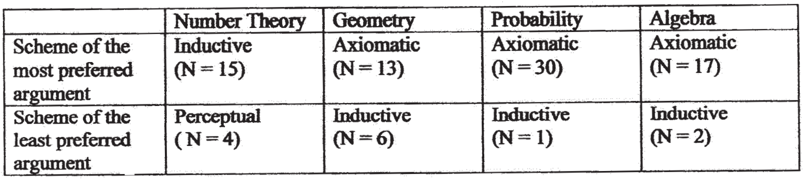
	Number Theory		Geometry		Probability		Algebra	
	Own	Preferred	Own	Preferred	Own	Preferred	Own	Preferred
NA	4	2	8	1	15	0	10	1
External	0		5		3		0	
Empirical	35	32	17	8	14	4	21	6
Analytical	2	7	11	32	9	37	10	34

However, students' tendency to move toward analytical schemes after examining the alternative arguments was clearly observed in all four content areas. The number of students who preferred an analytical scheme argument was significantly higher than the number of students who were able to produce them. Even in the case of the number theory problem, where the shift was less obvious as shown in Table 2, 10 students preferred the perceptual approaches shown in the argument compared to their own inductive approach. This hints the notion that students did tend to recognize and endorse more general mathematical explanations, even if they could not produce them themselves. This finding compels us to propose that if learners are inspired by arguments they can understand and find meaningful, there is great potential for their transition from one mode of reasoning to others.

Students' Preference for Argument Scheme

Table 3 illustrate the proof schemes of the most and the least preferred arguments by the participants.

Table 3.
Proof Schemes of the Most and Least Preferred Arguments in Each Problem



	Number Theory	Geometry	Probability	Algebra
Scheme of the most preferred argument	Inductive (N = 15)	Axiomatic (N = 13)	Axiomatic (N = 30)	Axiomatic (N = 17)
Scheme of the least preferred argument	Perceptual (N = 4)	Inductive (N = 6)	Inductive (N = 1)	Inductive (N = 2)

As the data indicated, the participants preferred inductive schemes in the number theory problem, while they favored the arguments with analytical schemes in the other three contexts. A closer inspection of data however revealed additional inconsistencies in the participants' choices. Among the 15 students who preferred the axiomatic proof of the geometry problem, only four considered the axiomatic proof of the number theory problem as *very convincing,* and only three of these four individuals chose this proof as their favorite option in the number theory problem. Of the 17 students who preferred the axiomatic proof of the algebra problem, only eight believed the axiomatic proof in the geometry problem was *very convincing,* and only five of the eight chose axiomatic proof as their favorite argument in the geometry problem. Of the 15 students who preferred the inductive proof in the number theory problem, only four considered this scheme proof in the geometry problem as *very convincing,* and only two of the four chose it as their favorite option in that context. This finding agreed with results of previous studies that suggested that the idea of proof and proving develops locally in a specific context instead of globally (Reid, 2011).

Comparing the reasoning schemes that students used across the four problems, significant differences were also detected among choices they had made. This difference was most prominent when comparing the participants' responses to the number theory problem with the ones they produced on the other three tasks. In particular, only two students adopted analytical approaches in the number theory problems, while about 10 students exhibited analytical reasoning in each of the other problems. This trend suggested that the transfer of reasoning skills from geometry to other mathematical fields, as expected by curriculum design, may not be automatic. That is, students may find it unnecessary to use analytical tools for proving statements outside the subject area of geometry and persist on relying on other schemes when reasoning accuracy of statements in different content areas.

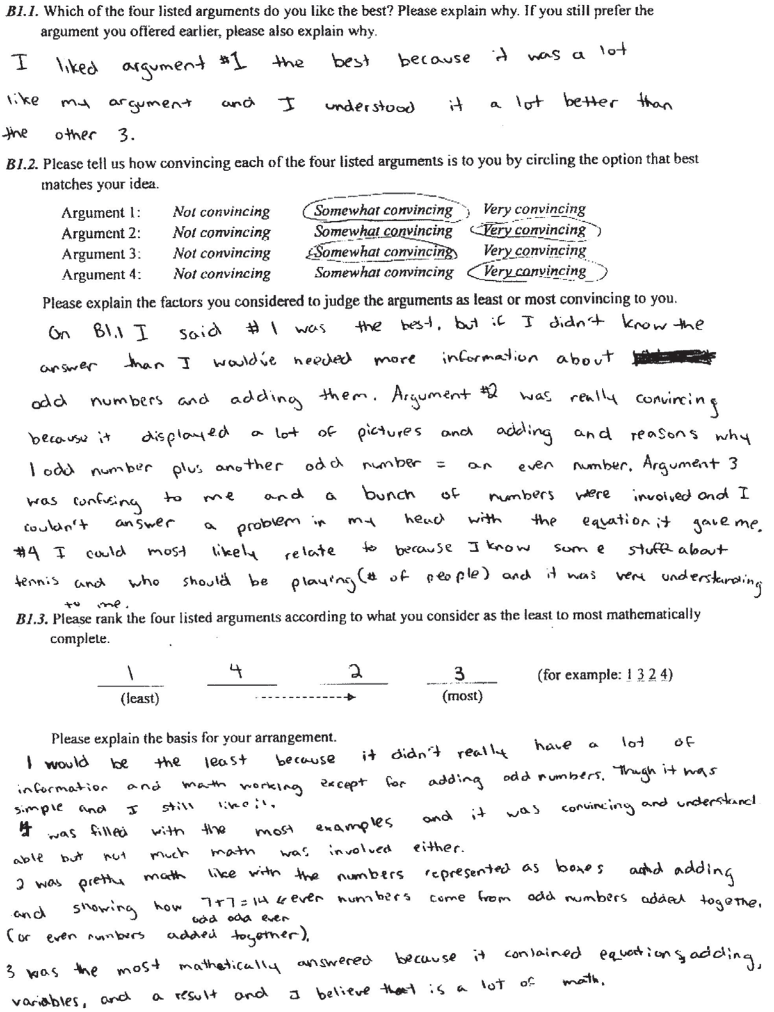
Shifts in Preference: Understanding the Rationale for Choice

To better understand the motive behind the shift in the students' preferences, we searched for any potential relationships that may have existed between the participants' judgment of *convincing* arguments and *mathematically complete* arguments. Furthermore, we examined whether students considered arguments with analytical proof schemes to be more convincing or mathematically complete than those with empirical proof schemes.

Preferred vs. convincing arguments. In conducting the analysis, we first examined whether the participants' preferred arguments were also convincing to them. Each participant offered a SJC score for each preferred argument. These preferred arguments were rated an average SJC score of 2.94 out of 3. This high score was not surprising; it seemed natural that students would prefer arguments that were *convincing* to them (Balacheff, 1991; Recio & Godino, 2001). Nonetheless, it was noted that in 10 cases students had stated that their preferred arguments were only *somewhat convincing* to them (the 10 cases were detected among seven students and three of these students each contributed two cases), implying that students did not always favour arguments that they found *convincing.* The explanations that the students provided confirmed that their choices were neither randomly nor carelessly made (Recio & Godino, 2001). An illustrative example of the tacit thinking behind the students' choices is presented in Figure 5. Note that in this explanation the participant clearly indicated that while she *"liked Argument 1 the best"* she had rated it as *somewhat convincing.* Although the student had ranked Arguments 2 and 4 as *very convincing,* she still favored Argument 1. Indeed, she identified this same argument as least *mathematically complete.* Such deliberate selection as typical of the entire group's rationale for the choices they had made.

Preferred vs. mathematically complete arguments. We further examined whether the participants' favorite arguments were also considered as mathematically complete by them. Each participant offered a SJMC score for each preferred argument. We calculated the average SJMC score of the arguments that students chose as their favorite and found these arguments were rated an average SJMC score of 3.73 out of 4. This result indicated that a majority of the participants preferred arguments that they considered as *mathematically complete.* However, we also detected in 14 cases that students proposed that their preferred arguments were either the least of the second least *mathematically complete* (these 14 cases were detected among 10 students and four of them each contributed two cases). An example of this type of reasoning is also depicted in Figure 5.

The high SJC and SJMC scores for students' favorite arguments indicated consistency among what they preferred, what they considered *convincing,* and what they considered *mathematically complete* for the majority of participants. However, cases observed with the remainder of the subjects



B1.1. Which of the four listed arguments do you like the best? Please explain why. If you still prefer the argument you offered earlier, please also explain why.

I liked argument #1 the best because it was a lot
like my argument and I understood it a lot better than
the other 3.

B1.2. Please tell us how convincing each of the four listed arguments is to you by circling the option that best matches your idea.

Argument 1: *Not convincing* [✓ *Somewhat convincing*] *Very convincing*
Argument 2: *Not convincing* *Somewhat convincing* [✓ *Very convincing*]
Argument 3: *Not convincing* [✓ *Somewhat convincing*] *Very convincing*
Argument 4: *Not convincing* *Somewhat convincing* [✓ *Very convincing*]

Please explain the factors you considered to judge the arguments as least or most convincing to you.

On B1.1 I said #1 was the best, but if I didn't know the
answer than I would've needed more information about [redacted]
odd numbers and adding them. Argument #2 was really convincing
because it displayed a lot of pictures and adding and reasons why
1 odd number plus another odd number = an even number. Argument 3
was confusing to me and a bunch of numbers were involved and I
couldn't answer a problem in my head with the equation it gave me.
#4 I could most likely relate to because I know some stuff about
tennis and who should be playing (# of people) and it was very understanding
to me.

B1.3. Please rank the four listed arguments according to what you consider as the least to most mathematically complete.

1 4 2 3 (for example: 1 3 2 4)
(least) ------------→ (most)

Please explain the basis for your arrangement.

have a lot of
1 would be the least because it didn't really
information and math working except for adding odd numbers. Though it was
simple and I still like it.
and it was convincing and understand
4 was filled with the most examples
able but not much math was involved either.
2 was pretty math like with the numbers represented as boxes and adding
and showing how 7+7=14 & even numbers come from odd numbers added togethe,
odd odd even
(or even numbers added together).
3 was the most mathetically answered because it contained equations, adding,
variables, and a result and I believe that is a lot of math.

Figure 5. An illustrative example of an eighth grader's responses to the *Survey of Reasoning*

revealed an inconsistency among their responses to the three questions. Such findings implied that neither the *convincingness* nor the *mathematical completeness* of an argument solely determined students' preference of arguments.

Empirical vs. analytical proof schemes. Our analysis in the previous section concerned the participants' preferential choices of *convincing* and *mathematically complete* arguments presented in each problem context. In extending this analysis, we examined if the differences in scores could be attributed to the arguments' particular scheme (empirical or analytical). All

16 alternative arguments on the *Survey of Reasoning* were divided into two groups based on their underlying schemes (see Table 1 and Figure 2). The empirical group and the analytical group each consisted of eight arguments.

We first computed the average SJC score for all arguments based on the entire sample's responses, as illustrated in Figure 6a. Data indicated that the participants, most prominently, considered the analytical arguments more convincing than the empirical arguments. This result agreed with findings of previous studies that posited students are aware of the limitations of the empirical arguments (Healy & Hoyles, 2000). Note that when considering only those students who believed an argument was *mathematically complete,* the between-group difference of SJC scores was largely reduced (see Figure 6b). The revised average SJC score for each argument was calculated based on the responses of those who had marked this argument as the most or second most *mathematically complete* in the follow-up question.

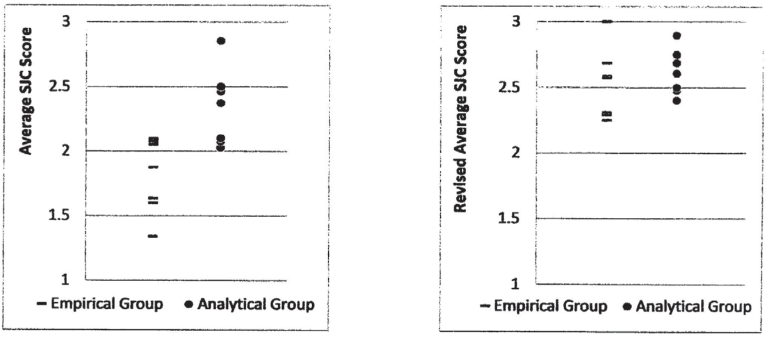


Figure 6a. Average SJC score of each argument type.

Figure 6b. Revised Average SJC score of each argument type.

Figure 6. SJC Scores

Schoenfeld (1988) and Chazan (1993) previously reported that students in their studies remained unconvinced of validity of statements even after generating proofs they had considered as *mathematical complete.* Our findings suggest a different pattern. That is, in our case, if students found an argument to be *mathematically complete,* they were likely to claim it was *convincing* as well, regardless of its reasoning scheme.

We further examined whether the students considered analytical arguments more *mathematically complete.* In doing so, we computed the average SJMC score for all the arguments of the entire sample, as illustrated in Figure 7a. As shown in the graph, the students generally considered the arguments with analytical proof schemes to be more *mathematically complete* than those with empirical proof schemes. We conjecture that aside from the content of the argument itself, the appearance of analytical arguments might

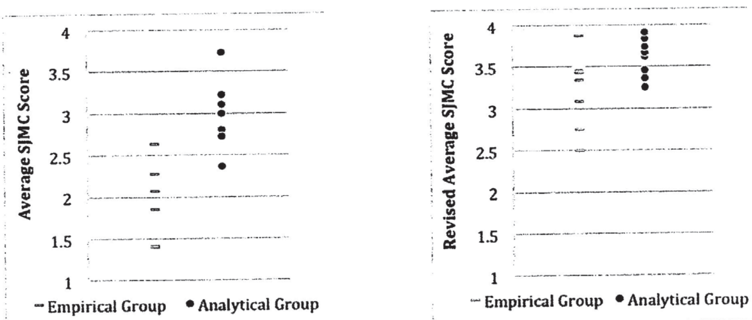


Figure 7a. Average SJMC score of each alternative argument.

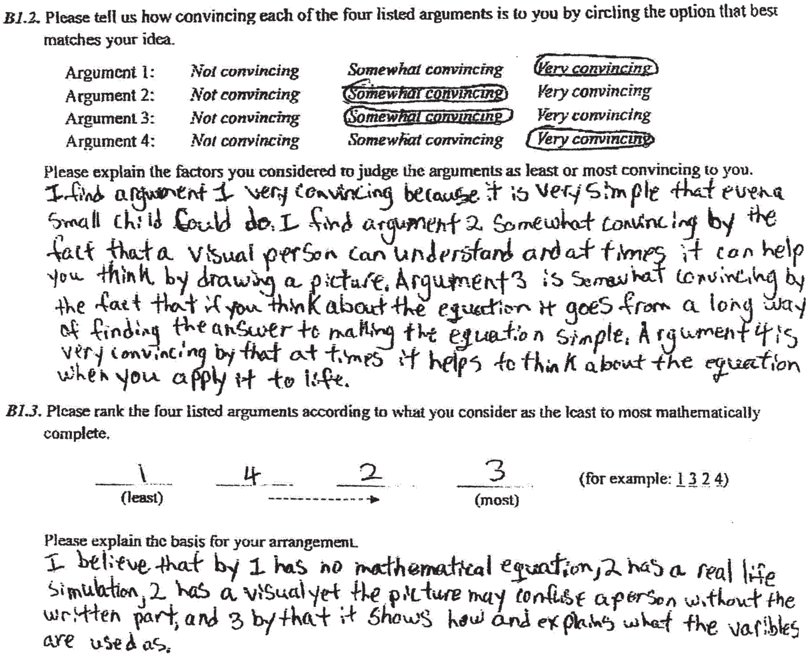
Figure 7b. Revised Average SJMC score of each alternative argument.

Figure 7.

have contributed to students' ranking as well (Harel & Sowder, 1998, 2007).

Next, we considered whether the arguments that the students had marked as *convincing* were also characterized as *mathematically complete.* In conducting the analysis, we computed the revised average SJMC score only based on data from students who considered this argument as *very convincing.* Results (see Figure 7b) showed that although the revised SJMC score increased for both groups, the students who were *convinced* by an analytical proof were still more likely to recognize the *mathematical completeness* of the corresponding proof, compared to those who were *convinced* by an empirically based proof. In other words, those students who had considered an empirically based argument as convincing did not simultaneously identify it as *mathematically complete.* While the participants worked with an analytical scheme proof, greater degree of consistency was observed between their judgment about the argument's *convincingness* and *mathematical completeness.* For instance, 12 students marked Argument 3 (axiomatic proof scheme) in the number theory problem was *very convincing.* Among the 12, 10 individuals labelled this argument as the most *mathematically complete,* and two ranked it as the second most *mathematically complete.* In the same context 14 students marked Argument 1 (inductive proof scheme) as *very convincing,* however three of them ranked it as the least of second least *mathematically complete* option. The explanation offered by one of the participants (see Figure 8) might shed light on this issue. As illustrated, she indicated that both Arguments 1 and 4 (inductive and perceptual proof scheme, respectively) were *very convincing* to her, however she marked them as the least and second least *mathematically complete* arguments, since neither one had matched her personal standards of simplicity and accessibility (Recio & Godino, 2001). When assessing the *mathematical completeness,* the student had considered appearance, form and indicators of the statement (Tall et al., 2012). The existence of a gap (Healy & Hoyles,

2000) between what was considered as *convincing* and what was perceived as *mathematically complete* was clearly demonstrated by our data.



B1.2. Please tell us how convincing each of the four listed arguments is to you by circling the option that best matches your idea.

Argument 1:	*Not convincing*	*Somewhat convincing*	[✓ *Very convincing*]
Argument 2:	*Not convincing*	[✓ *Somewhat convincing*]	*Very convincing*
Argument 3:	*Not convincing*	[✓ *Somewhat convincing*]	*Very convincing*
Argument 4:	*Not convincing*	*Somewhat convincing*	[✓ *Very convincing*]

Please explain the factors you considered to judge the arguments as least or most convincing to you.

I find argument 1 very convincing because it is very simple that even a small child Could do. I find argument 2 Somewhat convincing by the fact that a visual person can understand and at times it can help you think by drawing a picture. Argument 3 is Somewhat convincing by the fact that if you think about the equation it goes from a long way of finding the answer to making the equation simple. Argument 4 is very convincing by that at times it helps to think about the equation when you apply it to life.

B1.3. Please rank the four listed arguments according to what you consider as the least to most mathematically complete.

1	4	2	3	(for example: 1 3 2 4)
(least)	------------→		(most)	

Please explain the basis for your arrangement.

I believe that by 1 has no mathematical equation, 2 has a real life simulation, 2 has a visual yet the picture may confuse a person without the written part, and 3 by that it shows how and explains what the varibles are used as.

Figure 8. Typical explanation illustrative of the chasm

Discussion and Summary

Our findings offer that the participants in our studies adopted and determined their preferred reasoning schemes based on the concrete contexts with which they worked instead of following a broader uniform scheme when approaching tasks that elicited proving from them. This implies that the transfer of reasoning skills from one area (typically geometry) to other mathematical fields, as expected by current curriculum design, did not occur. When comparing the reasoning schemes that the students used in constructing their own mathematical explanations with those they preferred, we observed that the participants were not biased when judging their own methods for completeness or accuracy. Indeed, when confronted with alternative argument types they exhibited a tendency to favor those that involved analytical reasoning. The results suggest that a potentially productive pathway toward building children's proving capacity might be linked to their experiences with a wide range of mathematical contexts and not limited to geometry (Stylianides, 2007; Tall et al., 2012).

In this work, we also studied the relationship among the arguments that the students considered as *convincing* or *mathematically complete* and ways in which these judgements may have impacted their preference of the arguments. The participants considered the analytical arguments more *convincing* and more *mathematically complete* than the empirical arguments. Additionally, a gap was detected between what participants considered as *convincing* and those they marked as *mathematically complete.* In particular, participants in general considered *mathematically complete* arguments to be *convincing;* however, they expressed in quite a few cases that a *convincing* argument might not be *mathematically complete.* The second phenomenon was particularly evident when they were evaluating arguments with empirical proof schemes.

Our data indicated that students did not always equate *convincingness* with *mathematical completeness.* Such a perspective could impede the internalization of the process and the value of mathematical proof. Therefore, cases that illustrated an inconsistency between the students' perception of *convincingness* and *mathematical completeness,* though not prevailing, were of great concern since they might provide fertile ground for extending understanding of why students might have difficulties in learning proofs. We noticed from students' explanations that those who offered consistent ratings for the *convincingness* and *mathematical completeness* of an argument didn't necessarily understand the meaning of the two indicators. They could have simply guessed that they were expected to judge upon the same idea and then offer consistent responses. Therefore, the actual gap between the two conceptions of proof (Healy & Hoyles, 2000) could be more significant than what was depicted by our data. Improvement of the wording and setting of the survey and follow-up interviews might help better capture the gap, and we suggest a need to further examine the factors that motivate the gap. Up to this point, we observed that students offered different reasons to judge the *convincingness* and *mathematical completeness* of arguments. When explaining their selection of the *convincingness* levels, the students could have considered simplicity and accessibility. When assessing the *mathematical completeness,* the students might have focused on the appearance, form, and indicators. Further investigation of these reasons behind students' judgment would contribute to establishing an understanding of factors that might encourage or impede students' perception and appreciation of analytical proofs.

Reflections on the Proof Schemes Model

Harel and Sowder's (1998) theoretical construct of proof schemes served as a valuable analytical tool in our study since it allowed us to detect participants' choices and preferences for different types of arguments embedded in different tasks. However, our results suggest that a refinement of the model is needed so to accurately link the learners' choices of proof schemes to the mathematical content of the tasks that elicit reasoning. Understanding the intricacies of such a link enables us to better understand the connection between the reasoning schemes that learners use and the mathematical maturity that might be exhibited in their classes. These points are further explained below.

First, we found that arguments that seemingly embodied the same schemes elicited vastly different levels of mathematical maturity from the same individuals. The individuals' choices were closely linked to the topic in these arguments were used. Consider for instance the empirical proof scheme as an example, when a student checked a property for a few cases in a number theory problem (say the square of an even number must be divisible by 4) and suggested that this property would apply to all (infinitely many) cases, one could infer that the student might possess little awareness of or skills for generating formal mathematical arguments. However, if a student drew a triangle and demonstrated the validity of a conjecture about all triangles (say the formula of the sum of interior angles), this approach would be likely to be accepted as a valid proof to the standards of secondary school mathematics without asking the student to specify why the single triangle drawn could count for all others. In this situation it remains unclear whether the object that the student had chosen to work with was treated as a single case or a generic example (Balacheff, 1988). In presence of such ambiguity researchers might not be able to judge whether the generalization of a property to a broader domain may have occurred by the individual, particularly if the person fails to offer an elaboration using external representations accessible to the researchers. Indeed, it is possible that the generalization might have occurred but not expressed by the students. Even if the learner had offered evidence, researchers would rely on accepted conventions in different mathematical areas to decide whether representational choice was *mathematically complete.*

Second, within the same mathematical area standards for the judgment of reasoning schemes could still be different depending on the assumed premises. If a student represented an even number by "2n" and deduced that the square of an even number must be divisible by 4, it would usually be considered as a complete deductive mode of reasoning in a secondary mathematics classroom. However, those with an advanced algebra background could challenge this judgment and argue that such a representation (2n) might remain empirical. Another example could be found in the study of probability, where a student listed all the elements of the finite sample space, counted the cardinality of the event set, and obtained the probability of the event by division. Should this method be considered as an empirical approach? If the answer is yes, then how is this technique different from conducting a real experiment and claiming the answer from experimental results? If the answer is no, then how much more skill did it require beyond listing the cases one by one? What would happen if there were infinitely many cases so the students couldn't list them all?

In order to answer these questions more accurately, we must be more specific about the standards for the judgment of proof schemes. On one hand, the standards must respect the feature and convention of mathematical topics, which calls for a localization of Harel & Sowder's work to establish content specific models. On the other hand, the standards must respect the problem solver's tool-box (Reid, 2011); what learners may have taken for granted or perceived as needed to be achieved when confronted with proving tasks. The absence of relevant assessment instruments that capture what students know at a particular point and ways that they may understand the goal of the task serves as a major obstacle in improving proving instruction in schools. Additionally, the community's own inclination toward either identifying gaps in reasoning or comparing proof reading or proof producing among novice and experts has provided little insight into learners of sense making process, as it pertains to proving and reasoning. Meeting the challenge of making mathematical proving skills accessible to all children may indeed demand that researchers adopt new lenses and explore learners' potentials as opposed to classifying their thinking according to particular conventional rules.

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, UK: Holdder & Stoughton.

Balacheff, N. (1991). The benefit and limits of social interaction: The case of mathematical proof. In A. Bishop, Mellin-Olsen, E. & van Dormolen, J. (Eds.), *Mathematical knowledge: Its growth through teaching* (pp. 175-192). Dordrecht, Netherlands: Kluwer.

Balaguer, M. (2008). Mathematical Platonism. In B. Gold and Simons, R. A. (Eds.), *Proof and other dilemmas: Mathematics and philosophy* (pp. 179-204).

Ball, D. L., & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, Martin, W. G., & Schifer, D. (Eds.), *A research companion to principles and standards for school mathematics* (pp. 27-44). Reston, VA: National Council of Teachers of Mathematics.

Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7*(1&2), 23-40.

Boero, P. (Ed.) (2007). *Theorems in school: From history, epistemology and cognition to classroom practice.* Rotterdam, Netherland: Sense Publisher.

Brown, J. R. (2008). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures* (2nd ed.). Routledge: New York.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics, 24*(4), 359-387.

Chazan, D., & Lueke, H. M. (2009). Relationships between disciplinary knowledge and school mathematics: Implications for understanding the place of reasoning and proof in school mathematics. In D. A. Stylianou, Blanton, M. L., & Knuth, E. J. (Eds), *Teaching and learning proof across the grades: A K-16 perspective (*pp. 21-39). New York: Routledge.

Council of Chief State School Officers (2010). Common Core State Standards (Mathematics). National Governors Association Center for Best Practices. Washington, D.C.

Dreyfus, T. (2006). Linking theories in mathematics education. In A. Simpson (Ed.), *Retirement as process and concept: A festschrift for Eddie Gray and David Tall* (pp. 77-82). Prague, Czec Republic: Karlova Univerzita v Praze.

de Villiers, M. (2003). *Rethinking proof with the Geometer's Sketchpad.* Emeryville, CA: Key Curriculum Press.

Fawcett, H. P. (1995). *The nature of proof. Thirteenth Yearbook of the NCTM.* New York: Teachers College, Columbia University. (Original work published 1938).

González, G., & Herbst, P. (2006). Competing arguments for the geometry course: Why were American high school students to study geometry in the twentieth century? *International Journal for the History of Mathematics Education, 1*(1), 7-33.

Hanna, G., & de Villiers, M. (Eds.) (2012). *Proof and proving in mathematics education.* New York: Springer.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, Kaput, J., & Dubinsky, E. (Eds.), *Research in collegiate mathematics education III* (pp. 234-383). American Mathematical Society.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* Reston, VA: NCTM.

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

Heinze, A., & Reiss, K. (2009). Development argumentation and proof competencies in the mathematics classroom. In D. A. Stylianou, Blanton, M. L., & Knuth, E. J. (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 191-203). New York: Routledge.

Herbst, P., & Brach, C. (2006). Proving and doing proofs in high school geometry classes: What is it that is going on for students? *Cognition and Instruction, 24*(1), 73-122.

Jaffe, A. & Quinn, F. (1993). Theoretical mathematics: Toward a cultural synthesis of mathematics and theoretical physics. *Bulletin of the American Mathematics Society, 29,*1-13.

Krantz, S. (2007). The history and concept of mathematical proof. Retrieved from http://www.math.wustl.edu/~sk/eolss.pdf

Kuchemann, D., & Hoyles, C. (2009). From empirical to structural reasoning in mathematics: Tracking changes over time. In D. A. Stylianou, Blanton, M. L., & Knuth E. J. (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 171-190). New York: Routledge.

Lakatos, I. (1976). *Proofs and refutations.* Cambridge, UK: Cambridge University Press.

Liu, Y., & Manouchehri, A. (2012). What kinds of arguments do eighth graders prefer: Preliminary results from an exploratory study. In Van Zoest, L. R., Lo, J. J., & Kratkey, J. L., (Eds.), *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 226-229). Kalamazoo, MI: Western Michigan University.

Mason, J. (2009). Mathematics education: Theory, practice and memories over 50 years. In S. Lerman and B. Davids (Eds.), *Mathematical action & structures of noticing: Studies on John Mason's contribution to mathematics education* (pp. 1-4). Rotterdam: Sense Publisher.

Marrades, R., & Gutiérrez, A. (2000). Proofs produced by secondary students learning geometry in a dynamic computer environment. *Educational Studies in Mathematics, 44*(1&2), 87-125.

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

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics, 48*(1), 83-99.

Reid, D. A. (2011). Understanding proof and transforming teaching. In L. R. Wiest, & Lamberg, T. (Eds), *Proceedings of the 33rd annual meting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 15-30). Reno, NV: University of Nevada, Reno.

Schoenfeld, A. H. (2011). *How we think: A theory of goal-oriented decision making and its educational applications.* New York: Routledge.

Schoenfeld, A. H. (Ed.) (1994). *Mathematical thinking and problem solving.* Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disaster of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145-166.

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

Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (Eds.) (2009). *Teaching and learning proof across the grades: A K-16 Perspective.* Routledge: New York.

Tall, D., Yevdokimov, O., Koichu, B., Whiteley, W., Kondratieva, M., & Cheng, Y. (2012). Cognitive development of proof. In G. Hanna, & de Villiers, M. (Eds.), *Proof and proving in mathematics education* (pp. 13-50). New York: Springer.

Warning, S. (2000). *Can you prove it? Developing concepts of proof in primary and secondary schools.* Leicester, UK: The Mathematical Association.

Weber, K. (2001). Student difficult is constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics, 48*(1), 101-119.