LOOK ACADEMIC PUBLISHERS
OPEN ACCESS

IEJME — MATHEMATICS EDUCATION
2016, VOL. 11, NO. 7, 2373-2401

# What Do Eighth Grade Students Look for When Determining If a Mathematical Argument Is Convincing

Yating Liu[a], Jenna Tague[b], and Ravi Somayajulu[b]

[a]Old Dominion University, Norfolk, USA; [b]California State University, Fresno, USA

ABSTRACT

Existing research have found that students' creation and evaluation of mathematical proofs was inconsistent across content areas. Investigation into an explanation of the phenomena requires an analysis of students' thinking processes when they conduct an evaluation of mathematical arguments. This study is conceptualized to contribute to this investigation. The analysis investigated the aspects and features of arguments that impacted students' evaluation of the arguments. Eight 8th grade students participated in the interviews where they were asked to explain their rationale in evaluating arguments that justify conjectures from multiple strands of school mathematics. Interview data was coded using the Classification of Mathematical Argument (CMA) framework to identify the aspects and features of arguments that impacted students' evaluation of the arguments. A detailed analysis of each subject's interview response documented the complexity of each individual's rationale and offered descriptions of the various differences among individuals. Despite such individual differences, the study also revealed a common theme among the subjects in their reasoning, i.e. the accepted statements in an argument, instead of its mode of presentation or mode of argumentation, had the largest impact on the subjects' evaluation of an argument.

KEYWORDS
Argument evaluation, proof and reasoning, student thinking

ARTICLE HISTORY
Received 18 January 2016
Revised 20 April 2016
Accepted 24 April 2016

## Introduction

Nurturing students' mathematical reasoning and proving capacity has been recognized as fundamental aspects of mathematics education (CCSSO, 2010; NCTM, 2000). However, it is also well documented that students' understanding of mathematical proofs and their ability to conduct rigorous mathematical reasoning remains underdeveloped at all grade levels (Chazan & Lueke, 2009; Dreyfus, 1999; Harel & Sowder, 1998, 2007; Kuchemann & Hoyles, 2009; Weber, 2001).

CORRESPONDENCE Yating Liu y1liu@odu.edu

© 2016 Liu et al. Open Access terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/) apply. The license permits unrestricted use, distribution, and reproduction in any medium, on the condition that users give exact credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if they made any changes.

Usiskin (1982) conducted a large-scale study of secondary students' geometric reasoning levels using the van Hiele model (van Hiele, 1980; 1986). Results of the study indicated that "Few students (barely a quarter of the population at most) are at levels 4 or 5, the levels at which, according to the van Hiele theory, students are able to understand proof" (p. 40). Thirty years later, results from National Assessment of Educational Progress mathematics studies revealed similar results (NAEP, 2010). Only about a quarter of the Grade 12 participants (23% in 2005 and 26% in 2009) were at or above *Proficient* Level, where students are "able to test and validate geometric and algebraic conjectures using a variety of methods, including deductive reasoning and counterexamples" (p. 32); while very few of the participants (2% in 2005 and 3% in 2009) were at Advanced Level, where students are able to "provide mathematical justifications for their solutions, and make generalizations and provide mathematical justifications for those generalizations," "reflect on their reasoning," and "understand the role of hypotheses, deductive reasoning, and conclusions in geometric proofs and algebraic arguments made by themselves and others" (p. 32).

It is recognized that this failure might be due to the fact that proofs and the proving process are often taught as an isolated topic in a geometry course instead of as a conceptual tool for reasoning throughout the curriculum (Herbst & Brach, 2006; Reid, 2011). As a consequence, students tend to view proof as a special format of written work (e.g. two-column proof) instead of a dependable way to produce reliable arguments (Chazan, 1993; González & Herbst, 2006; Healy & Hoyles, 2000; Schoenfeld, 1988). To address the issue, recent reform efforts in the mathematics curriculum tended to place less emphasis on the format of proof while paying more attention to nurturing students' proof skills through the understanding of specific topics throughout the grades (de Villiers, 1990, 2003; Hanna, 2000a, 2000b; Reid, 2011). This is consistent with the call from *Principles and Standards for School Mathematics* (NCTM, 2000), which states that "reasoning and proof cannot simply be taught in a single unit on logic, for example, or by 'doing proofs' in geometry" (p. 56). Such a perspective situates proof as a mathematical method developing naturally through mathematical inquiry – the need for proof emerges when the need to explore, verify, and systemize mathematical ideas is recognized (De Villiers, 1990; Lakatos, 1976). To implement such an approach, it is not only important to carefully examine the development of content structure, but also to understand the nature of students' thinking in proof-related activities (Mejia-Ramos & Inglis, 2009).

One prominent work in understanding the students' mathematical reasoning is the proof scheme framework proposed by Harel and Sowders (1998). Extending previous research such as Bell (1976) and Balacheff (1988; 1991), Harel and Sowder organized the types of arguments students (primarily college mathematics majors) might use in various branches of mathematics and proposed a framework of proof schemes consisting of three main categories, i.e. "*external*," "*empirical*," & "*analytical*," each of which encompasses several subcategories. *External* proof schemes include instances where students determine the validity of an argument by referring to external sources, such as the appearance or authorship of the argument instead of its content. *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 that include resting upon definitions, postulates, or previously proven conjectures.

Harel and Sowder's (1998) proof scheme framework is powerful in categorizing the type of proof adopted by students, yet it does not serve the purpose of explaining *why* a particular proof scheme is used by an individual. Healy and Hoyles' (2000) study examined the factors that influence students' decisions on what type of proof was valid. Focusing on high-attaining 14- and 15-year-old students, Healy and Hoyles (2000) found that students' judgment was impacted by their understanding of the purpose of the proof (i.e. to satisfy the teacher or to convince themselves, also see de Villiers, 1990; 2003), their mathematical competence, the instruction they received (also see Hoyles, 1997; Herbst & Branch, 2006), and their genders.

Both Harel and Sowder (1998) and Healy and Hoyles (2000) found that students' creation and evaluation of mathematical proofs was inconsistent across content areas. Students might value and create a deductive proof in one context and yet rely on empirical verification in another context. Such findings are consistent with existing developmental models of proof understanding (e.g. Waring, 2000; Yang & Lin, 2008; Tall et al., 2012), where the overarching understanding and consistent use of deductive reasoning is not achieved until learners reach the higher levels. However, the basis on which students rely on different schemes is still unexplained. There is a lack of explanation of why students may rely on, for instance, theorems to prove a geometric property, but are simultaneously fully convinced by checking a few cases in proving a number theory conjecture. Investigation into an explanation of the phenomena requires an analysis of students' thinking processes when they conduct an evaluation of mathematical arguments. This study is conceptualized to contribute to this investigation.

## Theoretical framework

The purpose of this study is to explore what students look for when determining if a mathematical argument is convincing. In order to do so, the Classification of Mathematical Argument (CMA) framework (see Figure 1) was conceptualized based on Stylianides and Stylianides' (2008a) identification of three characteristics of proofs, and Harel and Sowder's (1998) proof scheme model to categorize the aspects and features of mathematical arguments.

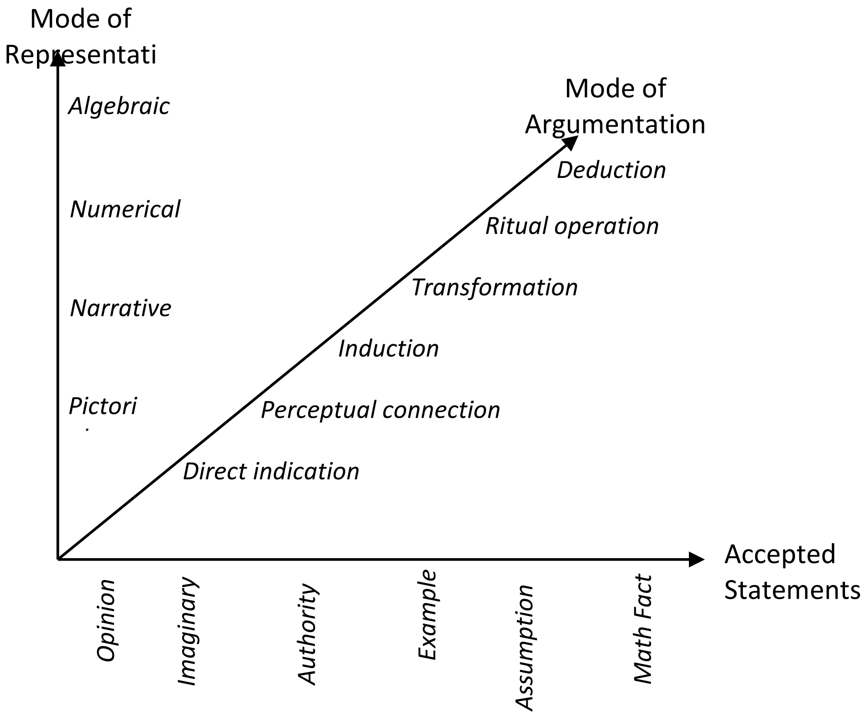



Figure 1. Classification of Mathematics Arguments (CMA) framework

Stylianides and Stylianides (2008a) identified three characteristics of mathematical proof: the mode of argument representations (which describes the format of expressions), the set of accepted statements (which states what is taken for granted), and the mode of argumentation (which represents the process of reasoning).

Stylianides and Stylianides (2008a) identified three modes of representations – *verbal*, *pictorial*, and *algebraic*. For this study, the *verbal* category was further differentiated as two categories, *narrative* and *numerical*. *Narrative* arguments refer to those using casual language without referring to exact numbers or using algebraic symbols. A typical example could be "Because the car is slower, it takes a longer time to get to the destination." *Numerical* arguments refer to those using Arabic digits and operational symbols (such as "+," "-," "<" and "( )"). For example, the argument "since 12 = 3 * 4, then 12 is a multiple of 3" is considered *numerical*. *Algebraic* arguments refer to those using the alphabet to represent mathematical concepts and communicate ideas. For instance, the argument "Since $x^2$-2x+1=$(x-1)^2$, then it must be non-negative" is considered *algebraic*. The last type is *pictorial* arguments, where pictures, figures, or other visual aids are provided to present concepts and to communicate ideas. An argument using the quantity of concrete figures to represent certain numbers is an example of *pictorial* arguments.

Different genres within the set of accepted statements and the mode of argumentation were informed through Harel and Sowder's (1998) proof scheme model. In particular, an accepted statement could be classified as *authority* (i.e.

what's stated by a respectful knowledge carrier, e.g. teachers, mathematicians, books, agreement of a community, etc.), *example* (i.e. result from an empirical test), *imaginary* (i.e. mental image created from recalling previous experience), *math fact* (i.e. a well known existing mathematical result), an *assumption* (i.e. promises assumed as the context of discussion), and *opinion* (conviction without an explicit reason).

The modes of argumentation used in this work include *direct indication*, *perceptual connection*, *induction*, *transformation* (Simon, 1996), *ritual operation* (Healy & Hoyles, 2000), and *deduction*. In *direct indication*, the conclusion is the required condition of source without any additional understanding (e.g. "Since the squares of a positive number, a negative number, and 0 are all non-negative, then the square of a real number is non-negative"). *Perceptual connection* refers to the linking of the source and the conclusion based on visualization or intuition. The argument "Since $f$(x) is a much longer expression than $g$(x), then $f$(x) must be larger" is an illustration of the use of *perceptual* connection in an argument. The use of metaphor/simile also falls into this category. *Induction* and *transformation* both refer to a conclusion informed by several pieces of empirical evidence; however, transformation involves a further investigation and noticing of properties that connect the empirical cases. For example, claiming that a property is true for all numbers purely based on testing a few numbers is considered *induction*; however, the use of *generic examples* (Balacheff, 1988), where certain general patterns were conceptualized during the study of the examples (e.g. envisioning the relationship between a person's distance from a street light and the length of this person's shadow), is considered *transformation*. *Ritual operation* and *deduction* both refer to a valid procedure; however, *ritual operation* refers to the use of a standardized procedure without knowing its limitations or why it works, while *deduction* refers to reasoning with an understanding of the logic between each of the steps in the process.

Throughout the rest of this paper, the mode of representation, the set of accepted statements, and the mode of argumentation will be called the "aspects" of an argument, while each category of an aspect is called a "feature" of an argument.

It is important to note that there is a certain degree of uncertainty when classifying a mathematical argument, depending on how explicitly the accepted statements are identified and how detailed the processes of argumentation are explained. Consider, as an example, the following argument: "Since 2+2=4, 2+4=6, 2+6=8, 4+6=10, then the sum of two even numbers must also be even." The representation of the argument is *numerical*. The accepted statements are the *examples* (results of several trials). The mode of argumentation appears to be *induction*, but might raise some questions. For instance, it is possible that when a student claims the argument to be valid, he/she might have gained insights from the trial results without explicitly expressing the discoveries. If that is the case, what convinced the student is no longer just an *induction*, but instead the *transformation* he/she made through the observation of the given examples. Therefore, it is important to acknowledge that the mode of representation, the set of accepted statements, and the mode of argumentation are not determined by the argument in its written format, but through a subject's interpretation of the argument.

Since individuals often have different interpretations of the same argument, the investigation of what features of an argument influence a student's evaluation of it can only be conducted with a thorough understanding of the student's interpretation and understanding of the argument. Therefore, clinical interviews were adopted as a methodology in order to collect in-depth qualitative data about students' thinking and reasoning (McConaughy, & Achenbach, 2001).

## Methods

### Participants

Clinical interviews were conducted with eight public school eighth grade students from a U.S. Midwestern state. Gender appropriate pseudonyms (Allen, Blake, Cindy, Deb, Emily, Fiona, Grace, and Heather) are used for the rest of the article. All of the subjects were enrolled in Algebra I or an equivalent class (Integrated 8$^{th}$ Grade Mathematics) at the time of the interviews. Two subjects (Allen and Grace) were taking Honors Algebra I. The interviews were recorded at the end of the spring semester, and so the subjects were close to finishing their coursework for the school year. There were 2 male and 6 female students. All subjects were native English speakers. It is not expected that the 8 subjects could generalize to students with all demographic and academic backgrounds. However, it is expected that the findings about the individuals could shed light on the nature of how students' thinking plays a role in evaluating mathematical arguments.

The choice of 8$^{th}$ graders as the subjects of the study is primarily based on two considerations. First, according to Piaget's (1985) *Intellectual Development Stages*, middle school students are at a critical cognitive phase where they can begin to engage in abstract and logical thinking. Therefore, how they learn to evaluate different arguments at this stage could potentially impact their reasoning skills and thinking habits in their later academic years. Second, according to the curriculum standards (NCTM, 2000; CCSSO, 2010), most 8$^{th}$ grade students should have obtained a basic understanding of numbers, shapes, probabilistic chance, algebraic expressions, simple propositions and properties, and should be able to see the connections between concepts and ideas. However, they may not have adopted abstract thinking or deductive ways of mathematical reasoning using conventional proving techniques and forms. Therefore, 8$^{th}$ grade serves as a bridge between middle and high school mathematics and the link between informal and more formal and abstract mathematical reasoning, and thus a focus on 8$^{th}$ graders' thinking and reasoning could provide valuable information on how to initiate the development of more rigorous mathematical reasoning in secondary mathematics.

### Interview Design

The According to a survey conducted by Mejia-Ramos and Inglis (2009), the majority of studies on mathematical proof are concerned with students' estimation, exploration, and justification of a mathematical conjecture, while few studies attend to students' comprehension or evaluation of a given proof. In addition, instruments that assess students' comprehension of proofs are also underdeveloped (Mejia-Ramos et al., 2012). To address this gap in the literature, the interview instrument was designed to explore students' evaluation of mathematical arguments in various contexts.

Five mathematical problems (A, B, C, D, and E) were chosen for the interview instrument (see Appendix I). The instrument design was informed by Healy and Hoyles' (2000) study where one conjecture was proposed in each problem followed by several arguments, which validated the conjecture in different ways. The problems were embedded in the main strands of school mathematics as described by the curriculum: number theory, geometry, probability, and algebra, which provided distinct contexts to investigate subjects' thinking. Conjectures in Problems A, B, D, and E are true conjectures. The conjecture in Problem C is false, but the counterexamples to this conjecture are not obvious. The arguments in each problem were provided to validate the corresponding conjecture. Table 1 provides an overview of the mathematical content and a key feature of the arguments, as identified by the researcher, associated with each of the conjectures.

**Table 1.** Content of each conjecture and a key feature of each argument

| Problem A (number theory) | Problem B (geometry) | Problem C (geometry) | Problem D (algebra) | Problem E (probability) |
|---|---|---|---|---|
| A1 (inductive) | B1 (inductive) | C1 (inductive) | D1 (inductive) | E1 (inductive) |
| A2 (algebraic) | B2 (perceptual) | C2 (algebraic) | D2 (algebraic) | E2 (pictorial) |
| A3 (perceptual) | B3 (algebraic) | C3 (pictorial) | D3 (perceptual) | E3 (perceptual) |
| A4 (pictorial) | B4 (pictorial) | C4 (perceptual) | D4 (pictorial) | E4 (algebraic) |

The problems were presented to subjects on colored paper cards each. Each card had the size of 20cm by 5 cm and was printed with either a conjecture or an argument. A conjecture and the arguments about the conjecture were printed on the cards of the same color. During the interview, the subject first selected a problem (a bundle of the conjecture and the arguments) by a color of his or her choice. The subject was asked to separate and read the conjecture and arguments, and then place the arguments in a column from the most convincing (on the top of the column) to the least convincing (on the bottom of the column). The subject was then asked to explain his/her rationale of the ranking by explicitly comparing one argument to all of the other arguments within that problem. After the explanation, the subject repeated the same process with a different problem.

Throughout the interviews, the subjects were encouraged to explain their thoughts and were provided ample time and opportunities to do so. In cases when the subject encountered difficulty in explaining his/her thoughts, the interviewer relied on several follow-up questions to facilitate the discussion. Such questions included:

- Do you think that one of the arguments is wrong?
- Do you think that this argument showed the conjecture is always true without any exception?
- Does any argument help you understand the problem better? Why?
- Do you think that this argument's evidence cannot support its conclusion?

The columns of arguments were placed side-by-side so the subject could have an overview of what argument was ranked highly or lowly in every problem. (See Figure 2 for a sample ranking by a subject. Each card had the actual conjecture or argument instead of just a label as shown in the figure.)

| | | | | |
|---|---|---|---|---|
| Conjecture B | Conjecture C | Conjecture A | Conjecture D | Conjecture E |
| Argument B2 | Argument C2 | Argument A1 | Argument D3 | Argument E2 |
| Argument B1 | Argument C3 | Argument A3 | Argument D4 | Argument E1 |
| Argument B3 | Argument C4 | Argument A2 | Argument D1 | Argument E4 |
| Argument B4 | Argument C1 | Argument A4 | Argument D2 | Argument E3 |

**Figure 2.** Illustration of subject's view of their rankings of the arguments in all problems

Besides explaining his/her ranking of arguments within each problem, the subject was also asked to review his/her ranking across the problems and to explain if there were any general ideas that guided their rankings across the problems. The comparison focused on why the subject ranked the arguments with a similar feature (e.g. pictorial demonstration, use of algebra, use of specific examples, and etc.) consistently highly across the problems, consistently lowly across the problems, or inconsistently across the problems. This comparison allowed the subject to explicitly explain his/her view of a certain feature of the arguments and how this feature may impact his/her judgment, and hence allowed an analysis of the explanation based on the CMA framework.

## *Data Analysis*

The interview responses were transcribed verbatim and analyzed qualitatively. First, each comment made by a subject toward an argument was classified according to whether it referred to the mode of representation, the set of accepted statements, or the mode of argumentation. Second, the occurrence of comments on each of the aspects was counted to describe the subject's attention while engaged in argument evaluation. Third, the occurrences of comments on a specific feature of an argument were counted, and whether such feature had a positive or negative impact on the subject's evaluation of the argument was also documented. Lastly, subjects made some comments regarding factors that were not included in CMA framework (e.g. the length of the argument might have impacted some subjects' judgment). Those factors were categorized as non-CMA factors and were also studied to understand each subject's rationale in argument evaluation. Figure 3 illustrates the steps in the analysis.

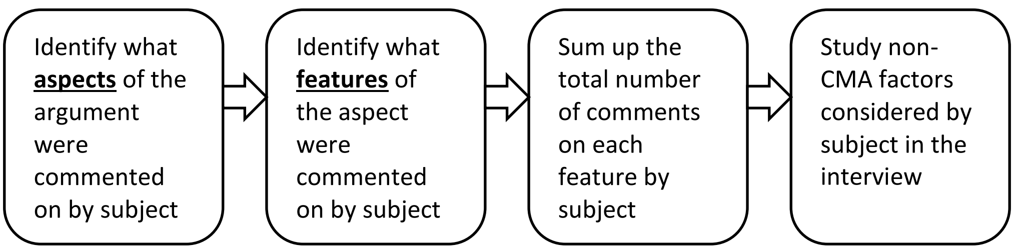



**Figure 3.** Steps in the interview analysis.

## Findings

The analysis of one subject's (Allen's) responses is provided to demonstrate the findings obtained from the study. The following section offered details about Allen's ranking of the arguments, his explanation of the ranking, and how his explanation was coded and interpreted to understand his rationale for evaluating mathematical arguments. The same analytical procedure was utilized for all the other seven subjects.

### *The case of Allen*

Allen was an 8th grade student enrolled in an Honors Algebra I class at the time of data collection. Figure 4 illustrates the rankings provided by Allen for each problem.

| Problem C | Problem B | Problem D | Problem A | Problem E |
|---|---|---|---|---|
| C2 (algebraic) | B3 (algebraic) | D4 (pictorial) | A4 (pictorial) | E2 (pictorial) |
| C4 (perceptual) | B4 (pictorial) | D1 (inductive) | A2 (algebraic) | E4 (algebraic) |
| C3 (pictorial) | B2 (perceptual) | D2 (algebraic) | A3 (perceptual) | E1 (inductive) |
| C1 (inductive) | B1 (inductive) | D3 (perceptual) | A1 (inductive) | E3 (perceptual) |

**Figure 4.** Argument rankings of provided by Allen

Allen's comments on the arguments were coded referring to the table of codes (see Table 2). Specifically, each of Allen's comments that referred to the mode of representation, accepted statements, or the mode of argumentation was coded with the corresponding capitalized letters, MR, AS, and MA, respectively, followed by a number that denotes the specific feature of that aspect.

**Table 2.** Table of codes

| Mode of Representation | | Accepted Statements | | Mode of Argumentation | |
|---|---|---|---|---|---|
| Pictorial: | MR1 | Authority: | AS1 | Direct: | MA1 |
| Narrative: | MR2 | Example: | AS2 | Perceptual: | MA2 |
| Numerical: | MR3 | Imaginary: | AS3 | Inductive: | MA3 |
| Algebraic: | MR4 | Math Fact: | AS4 | Transformational: | MA4 |
| | | Assumption: | AS5 | Ritual: | MA5 |
| | | Opinion: | AS6 | Deductive: | MA6 |

i). "P" denotes comments that didn't refer to the mode of representation, accepted statements, or mode of argumentation.
ii). "NA" denotes comments in which the subject claimed that he/she didn't understand the argument and didn't offer any explanation.
iii). A notation "-" was added behind the code to indicate that this feature made the argument less convincing to the subject.

The following clarifications are important in understanding the coding procedure.

1) Not all comments could be coded according to the CMA framework. In cases where the factors that contributed to the judgment were not identifiable or were not about the mode of representation, accepted statements, or mode of argumentation, the comment was coded "NC," denoting that there were non-

CMA factors that need to be further examined. It was denoted as having non-CMA factors since those reasons were not associated with any particular aspect of the argument. For example, the comment "it's not straightforward enough" was coded "NC" since it could apply to many different types of arguments. There were also cases when the subject indicated that he/she was not able to understand an argument. We used "NA" to denote such comments, suggesting that the subject was unable to provide an evaluation of the argument.

2) A certain feature could make an argument more or less convincing to the subject. To distinguish the effects different features had, a "-" was added to the end of a code if the identified feature made the argument less convincing to the subject.

3) A comment could refer to more than one feature or factor of an argument and hence could be multi-coded. For example, Allen made the comment that "it uses formulas which I know are fact, and I like seeing fact" in his explanation of why he considered Argument C2 valid. This comment was coded "AS4" and "MR4" since it was based on a mathematical fact as an accepted statement, which was expressed in an algebraic form.

4) There were scenarios in which it was difficult to judge what an argument meant based on the comment. In this case, the conversation before and after the comment was studied to determine the contextual meaning of the comment. For example, when reading the comment, "I'm not seeing very many supporting arguments," it was unclear what were the "supporting arguments" referred to by Allen. However, in reading the conversation that happened before the comment, where Allen talked about the need to see formulas in a convincing argument, it became clear that Allen was referring to mathematical facts as what he called "supporting arguments." Therefore, this comment was coded as "AS4."

The occurrence of each code was then summed and added into Figure 5 to describe the aspects and features of the arguments that influenced Allen's evaluation of their validity.

As shown in Figure 5, the total number of comments that focused on the mode of representation, accepted statements, and mode of argumentation were 27, 47, and 7, respectively, indicating that the accepted statements seemed to have the greatest impact on Allen's evaluation of the arguments. Among all types of accepted statements, Allen found that math facts (well known existing mathematical results) and examples (results from an immediate test) were reliable sources to establish an argument, each of which was referred to 17 and 18 times. His explanation was heavily rooted in the discussion of specific mathematical concepts (e.g. specific numbers' properties, specific geometric properties, meaning of graphs, etc.) instead of personal assumptions or opinions. This was highlighted by his claims that "when someone is trying to convince me of something, I would like facts" and "giving concrete numbers and facts and stating their observations of what they did the experiment on" would make an argument convincing. In addition, he clearly emphasized that "opinions and people doing things that I have not personally seen" did not make an argument valid to him. Similar statements were mentioned 8 times during the interview. Overall, Allen's comments demonstrated his need to see specific and concrete evidence in an argument in order to consider it valid.

| Total number of references to the mode of representation: **27** | | | | | | |
|---|---|---|---|---|---|---|
| | Pictorial | Narrative | Numerical | Algebraic | | |
| Positive | 12 | 1 | 1 | 12 | | |
| Negative | 0 | 1 | 0 | 0 | | |
| Total number of references to the accepted statements: **47** | | | | | | |
| | Authority | Example | Imaginary | Math Fact | Assumption | Opinion |
| Positive | 0 | 18 | 2 | 17 | 0 | 0 |
| Negative | 0 | 1 | 1 | 0 | 0 | 8 |
| Total number of references to the mode of argumentation: **7** | | | | | | |
| | Direct | Perceptual | Inductive | Transformational | Ritual | Deductive |
| Positive | 0 | 2 | 0 | 3 | 1 | 0 |
| Negative | 0 | 1 | 0 | 0 | 0 | 0 |

**Figure 5.** Aspects and features of arguments that influenced Allen's evaluation of their validity

The representation of arguments also influenced Allen's judgment. In particular, he indicated that pictorial and algebraic representations contributed to the validity of arguments. Each representation was referred to 12 times during the interview. Allen claimed that he loved "formulas, which are always in my mind second to visual representations." He also suggested that if "there's a combination of visual diagrams and formulas, that would be fabulous, that would be perfect." This tendency was backed up by his capability to represent variables with symbols and manipulate the symbols fluently, as well as the capability to connect graphs to the content of the problem.

Allen made fewer comments on the mode of argumentation. Among all the comments he made, only 7 referred to a certain way to connect the accepted statements to the conclusion of an argument. In 2, 3, and 1 case(s), respectively, Allen found a perceptual, transformational, and ritual reasoning valid. Allen was unable to recognize that showing a few examples would not prove a conjecture is always true. He considered an argument convincing "because it gives examples that worked."

In addition, Allen had personal standards that could not be captured by the CMA framework for deciding whether an argument was convincing. There were 14 comments that were coded as non-CMA factors, i.e. "NC." Nine of these comments concerned the simplicity of an argument, using terms such as "straightforward," "simple," and "quick" to explain why he was or was not convinced, while the other 6 comments referred to the clarity of the arguments (e.g. "There's always the showing, they're working it out"). These comments suggested that the pursuit of simplicity and clarity might sometimes override his preference on other aspects of an argument. For example, although Allen had repeatedly addressed the preference of seeing formulas, he claimed, in evaluating Argument D2, that "this is not straightforward… because it is a longer and more complicated and not straightforward enough formula" in explaining why he did not consider it convincing.

A clearer picture of Allen's rationale for evaluating mathematical arguments was formed when combining non-CMA factors and those characterized by the CMA framework. Allen viewed arguments that utilized precise descriptions and involved simple reasoning procedures as convincing. To him, known mathematical facts and concrete examples were the most straightforward accepted statements, while the pictorial and algebraic representations were the clearest ways to describe and relate those examples. However, since Allen was not yet able to reflect on the rigidness of logic embedded in an argument, the mode of argumentation was not among his major focuses. Arguments that used transformation, perceptual, and ritual reasoning might have been perceived as convincing by him. An argument was convincing to him as long as the reasoning looked "straightforward" to him, regardless of its logical rigidity.

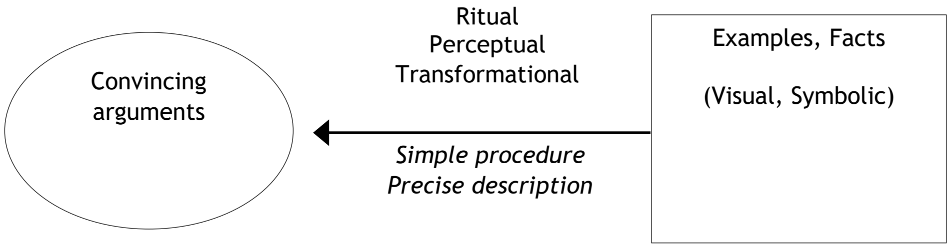



**Figure 6.** Illustration of Allen's rationale for evaluating mathematical arguments

With this platform, Allen's rankings of the arguments (see Figure 4) became more sensible. In Problem C, the clarity of the accepted statements provided in each argument determined their ranking. The accepted statements provided were ranked in the following order: C2 (the triangle area formula), C4 (imaginary triangle made by wire), C3 (drawn triangle within a transformation process), and C1 (a collection of triangles). Among these, the formula was the simplest and clear; the imaginary triangle made by wire was less clear, but also very simple; the triangle within a transformation process looked more complex; while the collection of triangles offered a mix of information and "trip[ped] [Allen] up for the first few seconds."

Arguments in Problem B were also ranked based on the simplicity and clarity of the evidence provided by them. Compared to his ranking for Problem C, the only difference was that the rankings of the visual and perceptual arguments were switched. Allen's explanation was that the image of the triangle made by wire was clearer to him than the image of a football field. Therefore, the argument based on the football field scene was less convincing to him.

In the other three arguments, Allen found the visual arguments to be the most convincing options while the algebraic arguments were ranked lower. A possible explanation was that in Problem C and B, both algebraic arguments contained well known mathematical facts (triangle area formula and the Pythagoras Theorem); however, in Problems D, A and E, the algebraic expressions were not well known formula or theorems but were used to represent the variables' relationships in the problem. Therefore, Allen's preference on algebraic expressions was not resolute but depended on the exact use of such expressions.

The different rankings of the inductive arguments across the problems could also be explained. Notice that in Problems A, B and C, the inductive arguments were considered the least convincing. This was because there was no actual example given in A1 and B1, while in C1, the examples seemed "confusing" to him. On the contrary, since D1 and E1 discussed more details about the examples, they were considered more convincing.

Overall, we found that the analysis of Allen's responses during the interview provided insights into the bases of his reasoning in determining the validity of different mathematical arguments. His reasoning was too complicated to be simply labeled as pro-algebra, pro-pictorial, or pro-induction. The pursuit of straightforward statements, the need to see mathematical facts and concrete examples as evidence, and preference towards visual and symbolic representation were all embedded in Allen's rationale in the evaluation of mathematical arguments, and all of them needed to be considered, in a specific context, to understand how Allen determined if an argument is convincing.

## Findings from all subjects

The Seven other subjects' interview data were analyzed using the same process as illustrated in Allen's case. Combining the data from each subject, it is clear that the subjects' view of which argument is convincing is highly distinct in every problem. Table 3 illustrates the differences (an argument received a score of 1, 2, 3 and 4, depending on the ranking by a subject, with 1 being the most convincing).

**Table 3.** Summary of the subjects' argument rankings

| | Allen | Blake | Cindy | Deb | Emily | Fiona | Grace | Heather |
|---|---|---|---|---|---|---|---|---|
| A1 | 4 | 1 | 3 | 2 | 4 | 4 | 4 | 2 |
| A2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 4 |
| A3 | 3 | 3 | 4 | 4 | 3 | 1 | 2 | 3 |
| A4 | 1 | 4 | 1 | 3 | 2 | 3 | 3 | 1 |
| B1 | 4 | 2 | 2 | 2 | 1 | 4 | 3 | 2 |
| B2 | 3 | 1 | 1 | 4 | 4 | 1 | 4 | 1 |
| B3 | 1 | 3 | 4 | 3 | 2 | 3 | 1 | 4 |
| B4 | 2 | 4 | 3 | 1 | 3 | 2 | 2 | 3 |
| C1 | 4 | 2 | 1 | 3 | 2 | 3 | 3 | 1 |
| C2 | 1 | 4 | 2 | 2 | 4 | 2 | 1 | 3 |
| C3 | 3 | 3 | 4 | 1 | 1 | 4 | 2 | 2 |
| C4 | 2 | 1 | 3 | 4 | 3 | 1 | 4 | 4 |
| D1 | 2 | 2 | 1 | 3 | 4 | 2 | 1 | 1 |
| D2 | 3 | 3 | 3 | 4 | 1 | 3 | 2 | 2 |
| D3 | 4 | 1 | 2 | 1 | 3 | 4 | 4 | 3 |
| D4 | 1 | 4 | 4 | 2 | 2 | 1 | 3 | 4 |
| E1 | 3 | 2 | 1 | 1 | 4 | 1 | 1 | 3 |
| E2 | 1 | 1 | 4 | 3 | 3 | 3 | 3 | 1 |
| E3 | 4 | 3 | 3 | 2 | 2 | 4 | 4 | 2 |
| E4 | 2 | 4 | 2 | 4 | 1 | 2 | 2 | 4 |

There is no clear pattern that any argument was considered convincing or not convincing by the subjects. Every argument (except for E3) was considered as the most convincing by some subjects while ranked as the least convincing by others. In other words, not a single argument could convince all the eight subjects that the corresponding conjecture is true. At the same time, any argument could convince at least one subject that the corresponding conjecture is true. Although no statistical conclusion on what argument is significantly more convincing than another one could be made due to the limited number of subjects, this finding suggested that it might be very unlikely to find a single most effective way to convince all students, and thus discussions from multiple perspectives using multiple representations might be a more promising way to convince learners with different thinking patterns.

The subjects' explanations were analyzed to investigate what factors might have impacted their judgment. Table 4 shows the number of references in each subject's explanation regarding each of the three aspects of arguments. The numbers in Table 4 suggest that the subjects paid the most attention to the accepted statements, followed by the mode of representation, with mode of argumentation being the least attended to among the three. Specifically, the accepted statements were the most referred aspect by 6 subjects, while the mode of argumentation was the least referred aspect by 7 subjects.

**Table 4.** Number of references regarding each aspect of arguments by the subjects

| | Allen | Blake | Cindy | Deb | Emily | Fiona | Grace | Heather | Total |
|---|---|---|---|---|---|---|---|---|---|
| Mode of Representation | 27 | 32 | 22 | 10 | 37 | 14 | 23 | 29 | 194 |
| Accepted Statements | 47 | 27 | 46 | 21 | 46 | 27 | 30 | 27 | 272 |
| Mode of Argumentation | 7 | 3 | 8 | 11 | 19 | 15 | 8 | 10 | 81 |

Table 5 further illustrates the subjects' perspective on what features of each aspect of the arguments contributed to their validity. The similarities and differences among the subjects are further specified in the following discussion.

## *Subjects' views on accepted statements*

The most prominent similarity among the subjects was that they all considered findings from testing a few examples as reliable accepted statements. This was observed in the comments from every subject on several if not all arguments.

Subjects' views on the use of mathematical facts were less consistent. Allen, Emily, and Grace indicated that they were likely to be convinced if an argument was based on a known mathematical fact. On the contrary, Blake seemed unwilling to use any established result and preferred exploring the problem from scratch. The other four subjects acknowledged that some known results (e.g. the triangle area formula) helped convince them an argument was true;

however, they only acknowledged these results as something they had heard about instead of as known established mathematical facts.

**Table 5.** Summary of the subjects' preferred features of arguments

| | Accepted Statement | Mode of Representation | Mode of Argumentation | Non-CMA factors |
|---|---|---|---|---|
| Allen | Examples, Math Facts | Pictorial, Algebraic | Transformational, Perceptual, Ritual | Simple procedure, Precise description |
| Blake | Examples, Imaginaries | Pictorial, Numerical, Narrative | Perceptual, Ritual | Easy to understand, Non-procedural |
| Cindy | Examples, Imaginaries | Pictorial, Numerical | Perceptual, Inductive | Easy to understand, Familiar procedure |
| Deb | Examples | Pictorial | Transformational | Easy to understand, Familiar procedure |
| Emily | Examples, Math Facts | Algebraic, Numerical | Deductive, Transformational | True for all cases |
| Fiona | Examples, Imaginaries | Pictorial, Narrative | Perceptual | Easy to understand, Relatable scenario |
| Grace | Examples, Math Facts | Algebraic, Numerical | Ritual, Perceptual, Transformational | Detailed procedure |
| Heather | Examples | Pictorial, Narrative, Numerical | Inductive, Ritual, Transformational | Easy to understand, Simple procedure |

The subjects' views of imaginaries (i.e. mental image created from recalling previous experience) also differed. To Blake, Cindy, and Fiona, imaginaries were a major source of evidence, while in Emily's view, an individual's brain can "skew everything" so imaginaries were definitely unreliable. To Allen, it depended on whether the imaginary was adequately clear to him.

Overall, the use of examples seemed to uniformly contribute to the subjects' evaluation of arguments, while each individual's view on the use of other sources, such as known mathematical facts and imaginaries, differed.

## *Subjects' views on mode of representation*

When looking at the mode of representation, pictorial representation was referenced the most and six subjects stated that visual aids could make an argument more convincing, especially when the image was simple and understandable to them. However, Emily and Grace expressed that they were unlikely to be convinced by pictorial arguments. Emily claimed that she was concerned that pictures and figures might misrepresent the problem, while

Grace believed pictorial illustration must be accompanied with narrative explanation in order to be persuasive.

Although not as commonly mentioned as pictorial representation, numerical representations were also often positively valued by the subjects. Five of the subjects believed a numerical representation makes an argument more convincing, and no subjects claimed that a numerical representation made an argument less convincing.

Narrative representation was the least commented type of representation. Some subjects demonstrated a higher need for narrative explanation than others. For example, Grace suggested that visual illustration alone was not convincing unless it was also accompanied by a narrative explanation. In contrast, Allen preferred to read equations and examine graphs and did not consider an argument convincing if it was too "wordy." Narrative representation could help the subjects to understand an argument. At the same time it could be difficult to use a narrative to describe some concepts or examples as precisely as using numerical, visual, or symbolic representations. Consequently, the subjects' evaluations of narrative descriptions depended highly on whether they understood the concepts embedded in narratives without seeing any specific numbers, images, or symbols, or whether they understood the numbers, images, or symbols in the absence of a narrative description.

Algebraic expressions were usually more abstract than ideas represented in the other three forms. Compared to the other three types of representations, the subjects showed the greatest differences in their views about algebraic representation. Students who understand the embedded ideas of algebra expressions often appreciate how clear and concise such expressions are in communicating ideas. For Emily, the algebraic representation could show the conjecture was true in every case. For Allen, the algebraic representation demonstrated the ideas clearly and concisely. For Grace, the algebraic representation helped her see the precise steps of the argument. Therefore, these three subjects found the algebraic representation positively contributed to their conviction. On the contrary, to those who had not yet adapted to algebraic representations, such arguments looked unintuitive and difficult, and hence were not convincing to them. For example, Blake considered algebraic terms confusing and not appropriate for his age group. Heather also found algebraically expressed theorems too abstract to communicate meaningful ideas. As a consequence, arguments using the algebraic representation were unconvincing to the two. The other three subjects neither claimed algebraic representations as helpful, nor did they find them confusing. Whether an argument was written by algebraic representations did not seem to contribute much to their evaluation of the mathematical arguments.

## *Subjects' views on mode of argumentation*

The mode of argumentation was the least commented aspect of arguments for almost every subject. Among the rare mentions of this aspect, Emily was also the only subject who found algebraic deduction the most reliable way to guarantee the conclusion of an argument to apply to general cases. In fact, she was the only subject who insisted that a convincing argument must show the conjecture was always true without any exception. According to the other subjects, this condition was not a requirement for a convincing argument.

Nonetheless, several subjects (Deb, Emily, Fiona, and Grace) articulated that showing a few examples might help them understand an argument, but were not sufficient to convince them that a conjecture was true. This suggested that some students were aware of the limitations of induction in proving general validity. Although they were not yet able to appreciate deductive reasoning, they had developed the ability to understand generic examples. For example, Deb could visualize that some geometric properties are stable when the shape was changing in particular ways. Allen could see that some values in an argument could be replaced by another value without violating the validity of each step in the argument. Overall, it was observed in five subjects' explanations that arguments that adopt transformational reasoning, and in particular, detecting and applying patterns from analyzing specific ideas, were considered convincing.

Perceptual connection was also applied by several subjects (including Allen, Blake, Cindy, Fiona, and Grace). Perceptual connection relates a given mathematical problem to imaginaries created from recalling previous experiences, and in many cases, such a connection was not precisely described, but was perceived by the subjects (e.g. by using a metaphor). Emily was the only subject who pointed out such a connection might not be a reliable way to build an argument.

Lastly, although ritual operations were rarely mentioned, they never contributed negatively to the subjects' evaluation of any argument in their explanations.

## Non-CMA factors

Non-CMA factors played an important role in the subjects' decision making and could be a major cause of the distinct evaluations of the same argument by different individuals.

Emily seemed to be the only person who believed a convincing argument should be one that proved the conjecture was always true. For the other subjects, this was not a guiding principle. This result is consistent with findings of existing research (e.g. Hersh, 2009; Selden, A., & Selden, 2003). Many subjects (Blake, Cindy, Deb, Fiona, and Heather) determined the credibility of an argument by examining how "easy" it was to understand it. However, there were differences in how they determined an argument is easy to understand. Blake found an argument easy to understand if it used easy language, easy examples, and easy pictorial illustrations. Cindy and Deb found an argument easy to understand if the concepts used in the argument and the steps in the reasoning procedure were familiar to them. Fiona considered an argument easy to understand only if the argument was built upon a real life scenario (as opposed to classroom experience) to which she could relate. Heather was able to appreciate more complex examples and pictorial demonstrations; however, she preferred an easy argument that did not involve a complex procedure (e.g. multiple steps).

Allen and Grace were the only two subjects who didn't claim that a convincing argument must be easy to understand. Allen claimed that he didn't have much difficulty understanding any argument used in the interview. Although he still personally preferred simple or "straightforward" arguments, he did not think whether an argument is easy to understand determines whether it is convincing. Similar to Allen, Grace also demonstrated an understanding of a wide range of arguments, but paid more attention to the details of arguments. Unlike Allen, Grace found that arguments with minimum wording often require readers to fill in the gaps of reasoning and hence are open to interpretation. She did not consider such argument to be convincing. For example, she did not consider pictorial illustrations alone to be convincing, since such illustrations must be accompanied with narrative explanations in order to avoid misinterpretation of their precise meanings.

### *Summary of the findings*

The analysis of the subjects' responses during the interviews revealed great differences among individuals in how they determined if an argument was convincing. Yet, an overall pattern was also observed: whether the subjects agreed with the accepted statements in an argument had the largest impact on their evaluation of the argument, followed by the mode of representation of the argument, while the mode of argumentation seemed to be the least considered aspect in their decision making.

Table 6 summarizes the similarities and differences in how the subjects determine if an argument is convincing. Consistent with existing research (e.g. Knuth, Choppin, & Bieda, 2009), when considering the accepted statements, the analysis of interview responses revealed that an argument based on empirical testing of examples was often considered convincing by the subjects. However, the subjects' views towards the use of mathematical facts and imaginaries differed.

In considering the mode of representation, the subjects often found numerical and narrative arguments easier to understand than algebraic ones. Pictorial illustrations could be helpful or confusing depending on the images or diagrams provided. Only one subject realized that the algebraic representation had the potential to prove the general validity of a conjecture. Some subjects found algebraic expressions concise and clear, while others viewed them as confusing and meaningless.

In considering the mode of argumentation, only one subject was aware that a valid argument must show the conclusion was always true without any exceptions. Half of the subjects realized argumentation based on induction was not reliable. Transformational and perceptual reasoning was widely viewed as convincing.

Lastly, several non-CMA factors were found to be a contributing factor to the subjects' evaluation of the arguments. The subjects' interview responses revealed that the perceived complexity of the arguments, students' familiarity with the contexts used in the arguments, and the clarity of the explanation presented seemed to have impacted the subjects' evaluation and judgment.

## Implication for research and practice

Similar to other studies based on the self-reflection of subjects, data obtained in this study had limitations in determining whether subjects' explanations actually reflected the rationale of their decisions (Dunning, Heath, & Suls, 2005). Additionally, the study used the number of subjects' comments on a certain feature of arguments as the indicator of whether the feature was an important factor in the subjects' decisions, which also involved a certain degree of bias since the topic of such comments was influenced by the flow of conversation occurring around when the subjects were making their judgment. Therefore, the value of the study can only be discussed with acknowledgement of these limitations.

**Table 6.** Similarities and differences in how the subjects determine if an argument is convincing

| | Similarities | Differences |
|---|---|---|
| Accepted Statements | • Findings based on testing a few examples were convincing.<br>• Authority, assumption and personal opinion were rarely referred to as convincing. | • Imaginaries and mathematical facts might or might not be viewed as reliable sources of evidence. |
| Mode of Representation | • Numerical and narrative arguments were usually easier to understand.<br>• Seeing a few numbers in an argument was helpful in most cases.<br>• Pictorial illustration was helpful if the provided image was understandable.<br>• Most subjects were not aware that algebraic representation denotes general cases. | • Pictorial illustration could be sufficient or not sufficient to demonstrate the validity of a conjecture.<br>• Narrative descriptions could be necessary or unnecessary.<br>• Algebraic expression could be concise and clear or confusing and meaningless. |
| Mode of Argumentation | • Deduction was rarely used or considered necessary.<br>• Transformation and perceptual connection was widely adopted.<br>• Ritual operation was rarely considered but was never unconvincing. | • Induction could be viewed as convincing, convincing in some situations, or not convincing at all. |
| Non-CMA factors | • Most subjects didn't focus on whether an argument could prove the conjecture was always true without any exception. | • Whether an argument was easy to understand was taken into consideration by some but not all subjects.<br>• Some subjects found arguments embedded in a familiar context more convincing.<br>• The subjects had different demands for the clarity of arguments. |

The existing trend of proof instruction continues to shift away from teaching students "the right way" of doing proofs and towards developing their abilities to generate arguments that can be used to convince oneself and others (Hanna & Jahnke, 1993; NCTM 2000; Healy & Hoyles, 2000; Stylianides & Stylianides, 2008b; Tall et al., 2012). Therefore, the process of nurturing mathematical reasoning should be built upon an understanding of how students convince themselves in the first place.

## *Reflection on what it means to develop mathematical reasoning "locally"*

The explanations provided by eight 8th grade students in the comparison of arguments within and across multiple contexts allowed researchers to gain insights of the bases of their reasoning. By coding students' explanations of how they evaluate the validity of mathematical arguments according to the CMA framework, results of the study suggested that the accepted statements of an argument had a greater influence on students' evaluation of the mathematical arguments than its mode of representation or mode of argumentation. Since the accepted statements are content specific while the mode of representation or mode of argumentation are more general characteristics, this finding is consistent with existing studies that suggest students develop an understanding of proof in local contexts (e.g. Freudenthal, 1971, 1973; Reid, 2011).

Results of the study suggested that helping students identify what accepted statements in a mathematically valid argument can be is an essential step in fostering their mathematical reasoning capacity. One finding of this study is that students' confidence in an argument was strongly influenced by the examination of concrete examples. This conclusion coincided with views of using examples and counterexamples to help students understand the construction of mathematical structures in a heuristic way (e.g. Lakatos, 1976; Knuth, Choppin, & Bieda, 2009; Stylianides & Stylianides, 2008b; von Glasersfeld, 1994). Although using examples to verify a statement is not a rigorous way to prove a statement, it does provide a concrete context for students to examine the mathematical concepts and procedures involved in the argument, and hence, to help them understand the problem better (Balacheff, 1988; de Villiers, 2003; Simon, 1996).

Results of the study also indicated that misunderstanding or rejection of mathematical facts often led to denial of mathematically valid arguments. Why can Side-Angle-Side imply congruency of triangles? Why does the distributive law hold for whole numbers, rational numbers, and real numbers? Why do you multiply the probability of each event to find out the likelihood of several independent events happening simultaneously? Related results and procedures are often memorized by students but the *whys* are often not investigated. Therefore, it is impossible for students to be fully convinced by arguments built upon these fundamental mathematical properties and results without a thorough understanding of such properties and results in the first place.

Since the validity of fundamental properties and results needs to be studied in a case-by-case manner, fostering proof capacity must be initiated in multiple strands of school mathematics. Students' understanding of the mathematical reasoning process grows concurrently with their experience in conducting such reasoning in different contexts. Only when their reasoning capacity within each context reaches certain levels are they able to identify features generally possessed by convincing arguments in these contexts. Therefore, results of the study suggest that it is more promising to develop mathematical reasoning based on an understanding of the fundamental properties and results within each mathematical content area, as opposed to authorizing a standard procedure that must be adopted in all areas of mathematics.

### *Reflection on the theoretical development of reasoning classification frameworks*

This study offered an explanation of why students' reliance on a certain type of argument is inconsistent across multiple contexts (Harel & Sowder, 1998; Healy & Hoyles, 2000; Author, 2013). It was evident in the results of this study, where those interviewed demonstrated different perceptions of the same type of arguments (e.g. inductive argument) in each problem. The status of pictorial representations' impact on students' conviction was also inconclusive. However, by analyzing the subjects' explanations, this study suggested that this inconsistency is caused by the mismatch of the criteria used by researchers to categorize the arguments and the factors considered by the subjects in evaluating an argument. For instance, a student may consider an inductive argument convincing in one context but another inductive argument not convincing in another context. While researchers may detect an inconsistency in the student's reasoning since both arguments are classified as "inductive," such inconsistency did not exist in the perspective of the student since he/she never noticed the common inductive feature of both arguments. Instead, the different evaluations might occur due to the fact that the student found the examples in one argument understandable, while the examples in the other argument were unfamiliar. So although it seemed that the student offered inconsistent views of whether inductive arguments are reliable, his/her need for familiar examples to verify the validity of an argument was consistent across content areas.

The development of argument classification models has often focused on two aspects of an argument, i.e. the mode of representation (e.g. algebraic vs. pictorial) and mode of argumentation (e.g. inductive vs. deductive). However, proof learners often pay more attention to the other aspect (i.e. the set of accepted statements of an argument). As such, students who haven't yet developed the ability to compare mathematical arguments across the content areas are unable to see the features researchers have used to label arguments. Instead, their evaluation of an argument was rooted in their understanding of its specific mathematical topic (e.g. whether they agree with the set of accepted statements).

With an emphasis on local development of mathematical reasoning ability, the absence of content specific proof/argument classification models becomes more critical. Considering the complexity of individual differences identified by this study, making any general conclusions to suggest certain kinds of arguments as more (or less) convincing to students is oversimplifying students' thinking patterns in argument evaluation. Current models measuring students' reasoning maturity or schemes are often based upon the synthesis of what was known about mathematical reasoning as a generalized method (e.g. Harel & Sowder, 1998; Simon, 1996; Tall et al., 2012; Waring, 2000). However, theories within specific content areas, especially areas other than geometry, remain underdeveloped. There are limited frameworks that synthesize how to make specific mathematical results convincing to students. Consequently, theories have not been built upon the features of local content and learners' understanding of such content. This is not to deny the existence of more general patterns in students' development of reasoning ability across the content areas. However, merely identifying these general patterns might not be sufficient to understand students' development of disciplinary reasoning skills and, as such, is limited in the quality of guidance it provides to support curricular instructional designs. Therefore, there is a critical need to develop content specific proof/argument classification and development models, which, perhaps, should also take some personal factors (e.g. the non-CMA factors identified in this study) into consideration.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**Yating Liu** holds a PhD in science education and now is an associate professor at Old Dominion University, Norfolk, USA.

**Jenna Tague** holds a PhD in science education and now is an associate professor at California State University, Fresno, USA.

**Ravi Somayajulu** holds a PhD in science education and now is an associate professor at California State University, Fresno, USA.

## 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.

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

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: AK-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.

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

de Villiers, M. (1998). An alternative approach to proof in dynamic geometry. In R. Lehrer & D. Chazan (Eds.), *New directions in teaching and learning geometry* (pp. 369-415). Lawrence Erlbaum.

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17–24.

Dunning, D., Heath, C., & Suls, J. M. (2005). Flawed self-assessment: Implications for education, and the workplace. *Psychological Science in the Public Interest, 5*, 69-106.

Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics, 38*(1/3), 85-109.

Freudenthal, H. (1971). Geometry between the devil and the deep sea. *Educational Studies in Mathematics, 3*(3-4), 413-435.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht: Reidel.

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. (2000a). A critical examination of three factors in the decline of proof. *Interchange, 31*(1), 21-33.

Hanna, G. (2000b). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, *44*, 5-23.

Hanna, G., & Jahnke, H. N. (Eds.). (1993). Aspects of proof [Special issue]. *Educational Studies in Mathematics, 24*(4).

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. 234-283). 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 in mathematics teaching and learning (pp. 805-842). Charlotte, NC: Information Age Publishing

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

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.

Hersh, R. (2009). What I would like my students to already know about proof. In D. A. Stylianou, Blanton, M. L., & Knuth E. J. (Eds.), *Teaching and Learning Proof Across the Grades: AK-16 Perspective* (pp. 17-20). New York: Routledge.

Hoyles, C. (1997). The curricular shaping of students' approaches to proof. *For the Learning of Mathematics, 17*(1),7-16.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge, UK: Cambridge University Press.

Author. (2013).

Knuth, E. J., Choppin, J. M., & Bieda, K. N. (2009). Middle school students' production of mathematical justifications. In D. A. Stylianou, Blanton, M. L., & Knuth, E. J. (Eds.), *Teaching and Learning Proof Across the Grades: AK-16 Perspective* (pp. 153-170). New York: Routledge.

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: AK-16 Perspective* (pp. 171-190). New York: Routledge.

Mejia-Ramos, J. P., & Inglis, M. (2009). Argumentative and proving activities in mathematics education research. In F.-L. Lin, F.-J. Hsieh, G. Hanna, and M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 88-93), Taipei, Taiwan.

Mejia-Ramos, J. P., Fuller, E., Weber, E.; Rhoads, K. & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*(1), 3-18.

McConaughy, S. H., & Achenbach, T. M. (2001). *Manual for the Semistructured Clinical Interview for Children and Adolescents* (2nd ed.). Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

National Assessment of Educational Progress (NAEP). (2010). *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. http://nces.ed.gov/nationsreportcard/pdf/main2009/2011455.pdf

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

Piaget, J. (1985). *The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development* (T. Brown & K. J. Thampy, Trans.). Chicago, IL: University of Chicago Press.

Reid, D. A. (2011). Understanding proof and transforming teaching. In L. R. Wiest, & Lamberg, T. (Eds.), *Proceedings of the 33rd Annual Meeting 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. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145–166.

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education, 34*(1), 4–36.

Simon, M. A. (1996). Beyond inductive and deductive reasoning: the search for a sense of knowing. *Educational Studies in Mathematics, 90*(2), 197-210.

Stylianides, G. J., & Stylianides, A. J. (2008a). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical thinking and learning, 10*(2), 103-133.

Stylianides, G. J., & Stylianides, A. J. (2008b). Enhancing undergraduate students' understanding of proof. *In Electronic Proceedings of the 11th Conference on Research in Undergraduate Mathematics Education* (http://sigmaa.maa.org/rume/crume2008/Proceedings/Stylianides&Stylianides_LONG(21).pdf), San Diego, CA.

Tall, D. et al. (2012). Cognitive development of proof. In G. Hanna, & de Villiers, M. (Eds.), *Proof and proving in mathematics education* (pp. 13-50). New York: Springer.

Van Hiele, P. M. (1980). Levels of thinking: How to meet them and how to avoid them. Paper presented at the *Research Presession to the annual meeting of the National Council of Teachers of Mathematics*. Seattle, WA.

van Hiele, P.M. (1986). Structure and insight: A theory of mathematics education. New York: Academic Press.

von Glasersfeld, E. (1994).A radical constructivist view of basic mathematical concepts. In Ernest, Paul (Ed.), *Constructing mathematical knowledge: Epistemology and mathematics education (Studies in mathematics education Vol. 4, pp.5-7).* Abingdon, Oxon: Routledge.

Usiskin, Z. (1982). *Van Hiele Levels and achievement in secondary school geometry*. The University of Chicago. Chicago, IL

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

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

Yang, K., & Lin, F. (2008). *A model of reading comprehension of geometry proof. Educational Studies in Mathematics, 67*(1), 59-76.

## Appendix I. Mathematics problems and arguments used in the interview

### PROBLEM A[1]

Shaina claimed that:

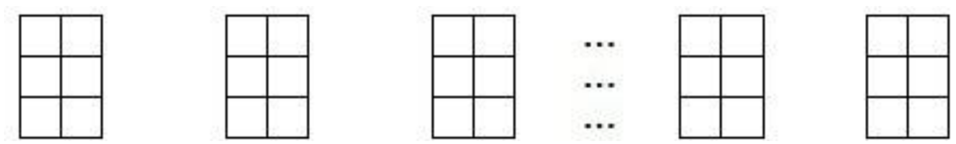
**"A multiple of 6 must also be a multiple of 3."**

**Argument A1**: I've tried plenty of multiples of 6 (like 12, 60, 606, etc.) and found they are multiples of 3 as well. So I am sure that Shaina's statement must be true.

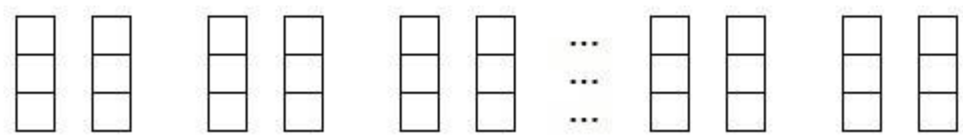
**Argument A2**: Any multiple of 6 can be written as $6n$. We know that $6n = 3 \cdot 2n$, which is a multiple of 3. Therefore a multiple of 6 must also be a multiple of 3.

**Argument A3**: If the total number of cookies is a multiple of 6, then we can put them into several boxes where each box contains 6 cookies. We can further divide each box into 2 packages, where each package contains 3 cookies. Now all the cookies are put into packages of 3. Therefore, the total amount of cookies must also be a multiple of 3.

**Argument A4**: The total number of square cards below is a multiple of 6:



We can rearrange the squares in this way:



Now we can see that a multiple of 6 must also be a multiple of 3.

[1] An item similar to Problem A was also used in Stylianides and Stylianides (2008b).

**PROBLEM B**

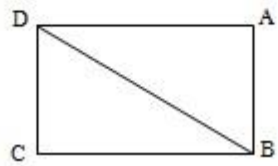
Ryan claimed that:

**"The diagonal of a rectangle must be longer than each of its sides."**

**Argument B1**: I've drawn several rectangles and measured the length of their sides and diagonals. I found that the diagonal of any of those rectangles is longer than any side of the same rectangle. So Ryan's statement must be true for all rectangles.

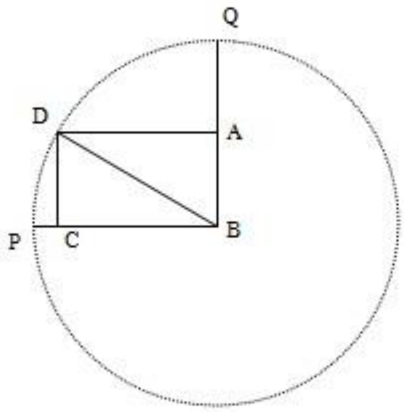
**Argument B2**: Imagine that you are standing on the corner of a football field. Then the diagonal of the field is definitely longer than any of its sides. So Ryan's claim must be right.

**Argument B3**: As shown in the figure below, ABCD is a rectangle.




Since ∠A = 90°, then by the Pythagorean Theorem,
BD^2 = AB^2 + AD^2.
So BD^2 > AB^2 and BD^2 > AD^2
(The notation X^2 means the square of X. For example, BD^2 means the square of BD). Therefore, BD is longer than AB and longer than AD.

**Argument B4**: Suppose ABCD is a rectangle. Draw a circle using B as the center and BD as the radius. From the figure shown, we can see that BD = BQ = BP. Since BC < BP and BA < BQ, then both BA and BC are shorter than BD. Therefore, the diagonal of a rectangle must be longer than any of its sides.

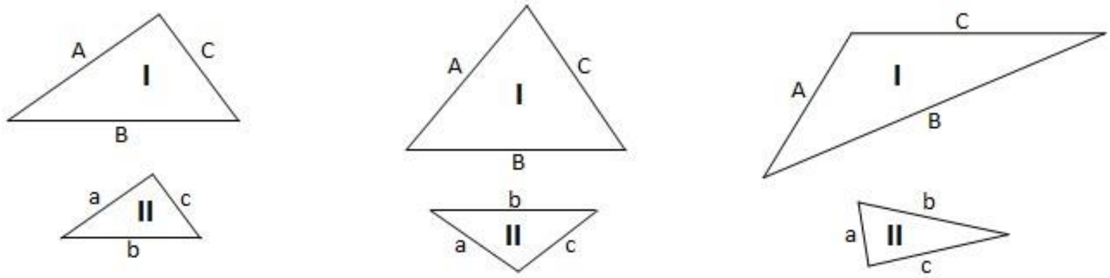



**PROBLEM C**

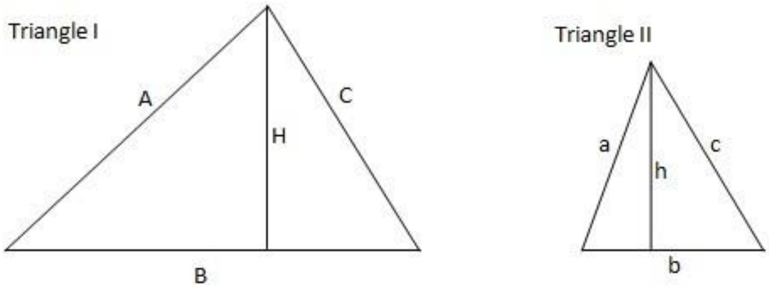
There are two triangles. The lengths of the three sides of Triangle I are $A$, $B$, and $C$ and the lengths of the three sides of Triangle II are $a$, $b$, and $c$. Jennifer claims that:

**"If $A > a$, $B > b$ and $C > c$, then the area of Triangle I must also be larger than Triangle II."**

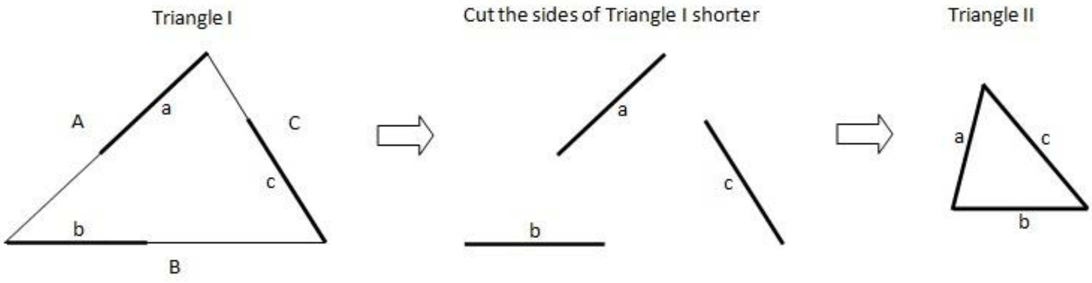
**Argument C1**: If $A = B = C = 2$, $a = b = c = 1$, then Triangle I is obviously larger than Triangle II. I also tried many other cases (as shown in the figures below) and found Triangle I always has an area larger than that of Triangle II. So I am sure Jennifer's claim must be correct.




**Argument C2**: We all know that the area of a triangle equals 1/2 of the product of its base and height. As shown in the figures below, the area of Triangle I = $BH/2$, and the area of Triangle II = $bh/2$. We know that $B > b$. In addition, since $A > a$ and $C > c$, then it must be true that $H > h$. So $BH/2$ must be larger than $bh/2$. Therefore, the area of Triangle I must be larger than the area of Triangle II.




**Argument C3**: As shown in the figures below, since each side of Triangle II is shorter than the corresponding side of Triangle I, we can cut each side of Triangle I shorter and then compose Triangle II using the shortened sides. Therefore, the area of Triangle II must be smaller than the area of Triangle I.




**Argument C4**: Since each side of Triangle I is longer than the corresponding side of Triangle II, then the perimeter of Triangle I must also be longer than the perimeter of Triangle II. If we make the two triangles using wires, then it needs a longer wire to make Triangle I than Triangle II. Using a longer wire we can make a larger triangle. Therefore, the area of Triangle I is definitely larger than the area of Triangle II.

**PROBLEM D**

The sales tax rate of the state where Ravi lives is 5%. Ravi is buying a new bike in a local bike store and has a $20 coupon[2]. Ravi claims that:

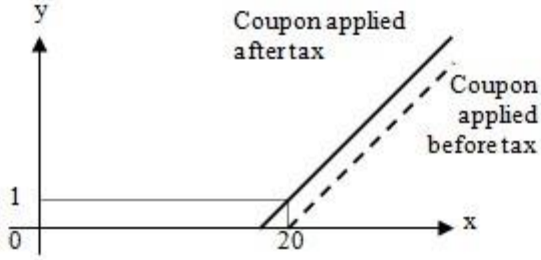
**"I can always save $1 if the $20 coupon is applied before tax rather than after tax, regardless of the actual price of the bike."**

**Argument D1**: Suppose the original price of the bike is $100.
If the coupon is applied before tax, then Ravi needs to pay
(100 – 20) × (1 + 5%) = 84 dollars.
If the coupon is applied after tax, then Ravi needs to pay
100 × (1 + 5%) – 20 = 85 dollars, which is $1 more than what he needs to pay if the coupon is applied before tax.
I tried some other possible prices of the bike, such as $200, $500, etc., and found he always pays $1 less if the coupon is applied before tax.
Therefore, I am sure Ravi's claim is always right.

**Argument D2**: Suppose the original price of the bike is $x$ dollars.
If the coupon is applied before tax, then Ravi needs to pay
$(x – 20) × (1 + 5\%) = 1.05x – 21$ dollars.
If the coupon is applied after tax, then Ravi needs to pay
$x × (1 + 5\%) – 20 = 1.05x – 20$ dollars.
Notice that $(1.05x – 20) – (1.05x – 21) = 1$. Therefore, Ravi always saves one more dollar if the coupon is applied before tax rather than after tax.

**Argument D3**: If the coupon is applied before tax, then Ravi doesn't need to pay the tax for the $20 discount. If the coupon is applied after tax, then he needs to pay the tax of the original price of the bike. Notice that $20 × 5% = 1. Therefore, Ravi always saves one more dollar if the coupon is applied before tax rather than after tax.

**Argument D4**: Let x be the original price of the bike and y be how much Ravi actually needs to pay (after applying the coupon and tax). Based on calculation, the graph below is generated by a graphing calculator to illustrate the two situations: the solid line represents how much Ravi needs to pay if the coupon is applied after tax; the dashed line represents how much he needs to pay if the coupon is applied before tax. From the




[2] The assumption that a bike costs more than $20 was not stated in the problem to test if subjects themselves might raise this question.

graph, we can see that the solid line is parallel to the dashed line and is always 1 unit above it. Therefore, Ravi can always save one more dollar if the coupon is applied before tax rather than after tax.

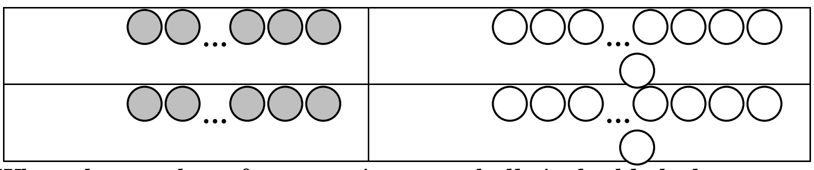
## PROBLEM E

There are some white and orange ping-pong balls in a box. You cannot see what's inside the box but you will get a reward if you pick out an orange ping-pong ball from the box. Jenna claims that:

**"If the number of white ping-pong balls and the number of orange ping-pong balls are both doubled, the chance for you to get a reward still stays the same."**

**Argument E1**: Suppose there are 2 orange ping-pong balls and 3 white ping-pong balls in the box, then the chance for you to get a reward is 2 out of 2+3, which is 40%. If the numbers of ping-pong balls of each color are both doubled, then there will be 4 orange ping-pong balls and 6 white ping-pong balls. Hence the chance for you to get a reward is 4 out of 4 + 6, which is also 40%. Therefore, the chance of winning the reward won't change.

**Argument E2**: As shown in the figure below, if the numbers of orange and white ping-pong balls are both doubled, the ratio between the ping-pong balls of the two colors will still be the same. Therefore, the chance of winning won't change.



**Argument E3**: When the number of orange ping-pong balls is doubled, the cases for winning the reward are also doubled. However, when the number of white ping-pong balls is doubled, the cases for not winning the reward are also doubled. As a result, the ratio of the cases of winning to the cases of not winning stays the same. Therefore, the chance of winning won't change.

**Argument E4**: Suppose there are $n$ orange ping-pong balls and $m$ white ping-pong balls in the box, then the chance for you to get a reward is $n / (n + m)$. If the numbers of ping-pong balls of each color are both doubled, then the chance for you to get a reward becomes $2n / (2n + 2m)$, which is equal to $n / (n + m)$. Therefore, the chance of winning the reward won't change.