Adolescent Reasoning in Mathematics: Exploring Middle School Students' Strategic Approaches in Empirical Justifications

Jennifer L. Cooper (jcooper4@wisc.edu) [A]
Candace A. Walkington (cwalkington@wisc.edu) [A]
Caroline C. Williams (ccwilliams3@wisc.edu) [C]
Olubukola A. Akinsiku (akinsiku@wisc.edu) [C]
Charles W. Kalish (cwkalish@wisc.edu) [B]
Amy B. Ellis (aellis1@wisc.edu) [C]
Eric J. Knuth (knuth@education.wisc.edu) [C]
[A] Wisconsin Center for Education Research, 1025 W. Johnson Street
[B] Department of Educational Psychology, 1025 W. Johnson Street
[C] Department of Curriculum and Instruction, 225 N. Mills Street
Madison, WI 53706 USA

Abstract

Twenty middle-school students participated in semi-structured interviews in which they were asked to assess the validity of two mathematical conjectures. In addition to being free to develop a valid proof as a justification, students were also asked to generate numeric examples to test the conjecture. Students demonstrated strategic reasoning in their empirical approaches by varying the quantity, parity, magnitude, and typicality of the numbers selected. These strategies were more developed in students who initially believed in the truth of the conjecture as well as in students who generated a valid, deductive proof. Emphasizing students' strategic selection of diverse examples parallels inductive reasoning in other domains. Strategic use of examples in justifying conjectures has the potential to assist students' development of deductive proof strategies.

Keywords: inductive reasoning; middle school mathematics; proof; empirical-based reasoning

Background

Many consider proof to be central to the discipline and practice of mathematics. Yet surprisingly, the role of proof in school mathematics has traditionally been peripheral at best, usually limited to high school geometry. More recently, however, mathematics educators and researchers are advocating that proof should play a central role in mathematics education. Reasoning about the properties, relationships, and patterns in math, as one does with proofs, supports the development of mathematical expertise.

Yet, despite the growing emphasis on justifying and proving in school mathematics, students rely overwhelmingly on examples to justify the truth of statements rather than using deductive proofs (e.g., Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009; Koedinger, 1998; Porteous, 1990). Many students fail to understand the nature of what counts as evidence and justification (Kloosterman & Lester, 2004). In mathematics, testing examples is not sufficient for proof – a deductive argument is necessary to cover all possible cases.

The preceding discussion regarding students' reliance on examples to "prove" the truth of statements (i.e., provide empirical-based justifications) is not meant to imply that examples do not play an important role in mathematical activity. Indeed, mathematicians often utilize examples to gain insight, develop an argument, and verify that an argument works (Alcock, 2004). The challenge remains, however, to help students learn to differentiate these appropriate uses of examples from their use as a primary means of justification.

Although reasoning inductively[1] features prominently in students' math justifications, the strategies underlying such reasoning are typically treated by mathematics educators as stumbling blocks to overcome rather than as objects of study in their own right or as starting points from which to foster the development of more sophisticated (deductive) ways of reasoning. The research has focused primarily on distinctions between the inductive, empirical approach and deductive justifications. Questions such as what might make one example or empirical justification stronger than another have not been well addressed.

In contrast, inductive strategies have been an ongoing focus of research in other domains such as biology where children and adults reason competently using inductive reasoning (e.g., Gelman & Kalish, 2006; Gopnik et al., 2004; Rhodes, Brickman, & Gelman, 2008). Inductive approaches and predictive inferences are appropriate in this domain, and they are supported by category knowledge. In particular, empirical justifications are rated as stronger when based on typical examples with high similarity to the category (Osherson et al., 1990). Having a diverse set of examples increases the coverage of the category. People's knowledge about the underlying category structure supports successful inferential reasoning (Osherson et al., 1990).

Effectively employing strategies to select informative examples depends, at least in part, on intuitions about similarity and typicality relations. It is unclear to what degree students have robust intuitions about the relations and category structure of mathematical objects and the

[1] Here we refer to making generalizations about a class of numbers based on observing or testing particular instances of that class, not mathematical induction, which is a valid method of proof.

extent to which such intuitions guide inductive inference. In an exploratory sorting task, middle school students used mathematically relevant features such as parity and factors to categorize the numbers (Knuth et al., 2009; in press). Dimensions such as these could underlie typicality ratings within a category.

Here, we focus on the qualities of examples chosen by middle school students to justify mathematical conjectures. The strategies underlying their inductive reasoning are explored by examining the typicality and the diversity of the example sets in order to gain insight into the mathematical knowledge underlying students' reasoning. However, as inductive strategies by themselves are insufficient as formal justifications in math, we also consider ways in which inductive strategies interact with the development of generalized, deductive proofs. Thus, this research initiates an in-depth investigation of empirical reasoning in mathematics that parallels research on inductive reasoning in other domains as it attempts to establish dimensions on which the strength of mathematical inductive reasoning can be rated.

Research Questions

In the current study, we address several questions. First, what approaches are employed by middle school students to evaluate the truth of conjectures? In answering this, we focus on a) the qualities of the examples chosen, b) the strategies that children use to select examples and c) the relationship between empirical approaches and valid proofs. Related to this is how initial reactions to the truth of the conjecture influence their subsequent reasoning.

Methods

We conducted semi-structured, videotaped interviews with 20 middle-school students (11 F, 9 M). The math grade levels reported (7 sixth-grade, 7 seventh-grade, and 5 eighth-grade or higher math courses) indicate the course year the student was currently in or had just completed as 7 students were in a math course above their year in school.

Each participant was asked to explore the validity of two mathematical conjectures (see Figure 1) during the first 20 minutes of the interview. The mathematical conjectures were selected to be statements for which proofs of different types would be accessible using middle school mathematics. First, participants were asked whether they believed each conjecture to be true for every number. The next questions asked participants "how they knew" their judgment of the truth and how they would figure it out. The researcher also asked the students to generate examples to test the conjectures. Once participants were convinced of the conjecture's truth, they were asked again to explain why the conjecture was always true and how they would show that to others.

The interviewer also asked participants to discuss the qualities of the examples they chose to test. Students were asked to classify the examples as typical or unusual and explain that classification. In addition, students were asked whether various pairs of numbers from their example set were similar or different. For each classification the student agreed to (typical or unusual; similar or different), the intentionality and valence of using those types of examples was assessed. Thus, the follow-up questions focused on whether the students' beliefs about typicality and diversity affected how they generated examples and their overall satisfaction with the approaches they used. Each student had multiple opportunities to explain and justify their reasoning as well as the opportunity to develop generalized proofs.

1) Whole Number Conjecture:
First, pick any whole number.
Second, add this number to the number before it and the number after it.
Your answer will always equal 3 times the number you started out with.

2) Even Number Conjecture:
First, pick any even number.
Second, add this number to half of itself.
Your answer will always be divisible by 3.

Figure 1: The mathematical conjectures presented.

Results and Discussion

Initial Reactions to Conjecture Truth

Half (47%) of the students were unwilling to specify an initial belief about the truth of the conjecture – they were either not sure or wanted to test the conjecture with a specific example. Of those students who did provide an initial reaction, the even conjecture was more frequently believed to be true (72%) than the whole number conjecture (40%). With the later analyses, we will see the subgroup of students who initially believed a conjecture to be true showed a different pattern of reasoning and justifications. The initial reaction was related to math experience. Students who had at least 8^{th} grade math were more likely to proceed directly to testing the conjecture (60% tested) than 6^{th} graders (7%) or 7^{th} graders (21%).

Types of Approaches

Overall, students' attempts to demonstrate the truth of a mathematical conjecture portrayed a diverse range of approaches including inductive reasoning through examples as well as deductively valid proof arguments. Justifications were coded according to Healy and Hoyles (2000) as being empirical, narrative, visual, or algebraic. Empirical justifications were based on the testing of specific examples. Narrative proofs explained why the property was true using verbal, deductive language. Visual proofs relied on drawings showing why the conjecture was true for a generic case (e.g., illustrating quantities being broken apart). Algebraic proofs used formal deductive statements of equality or equations. These latter three justification types reflect a more deductive approach in which a student demonstrates the validity of the conjecture for the general case; these three approaches will all be considered valid (i.e., non-empirical) proof strategies here. In contrast, empirical approaches do not fully justify the truth of the conjecture.

Empirical approaches were by far the most common strategy employed by participants (at least one example was tested in all but one case[2]), but it is important to remember that the interviewer explicitly asked the participants to generate examples to test.

Slightly under half of the conjectures were accompanied by valid proofs. Five students produced valid proofs for both conjectures; there were 8 students (evenly divided between the two conjectures) who produced only one valid proof, for a total of 18 proofs. Narrative proofs were the most frequent, followed by visual and algebraic proofs (see Figure 2). The probabilities of producing a valid proof or a particular type of proof were not affected by the particular conjecture.

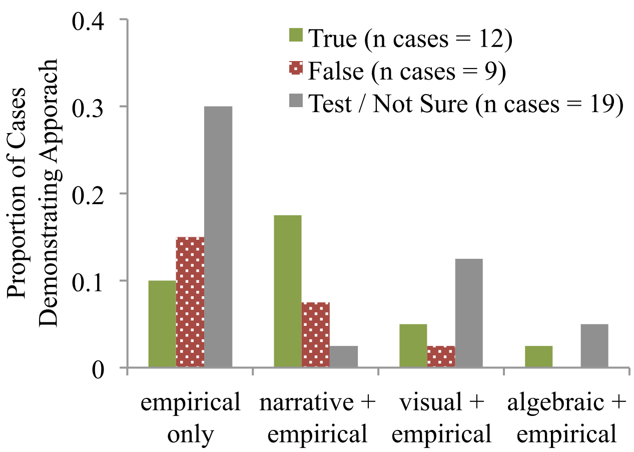

Figure 2: Empirical approaches predominated, but students' approaches[3] were affected by initial reaction.

The relatively higher frequencies of narrative proofs, particularly when the conjecture was believed to be true, resulted in valid proofs (i.e., non-empirical approaches) being more likely for subjects who initially believed the conjecture to be true (58% produced proofs) than other reactions (36% produced proofs). One possibility is that students are more inclined to generate valid proofs in order to support the veracity of their initial reaction. Empirical reasoning was the only approach used by half of the subjects, as is commonly found with this age range.

Empirical Justifications and Valid Proofs

Considering the order of the empirical versus valid proof approaches allows a more thorough assessment of the participants' overall level of competence with justifications and ways in which empirical and deductive approaches can be mutually supportive. Only cases on which a valid proof (n = 18 cases) was produced are considered here.

If students view proof as sufficient evidence to support a conjecture, one would expect the students' reasoning to end after generating a valid proof. While this was the case for the majority (78%) of the conjectures with a valid proof, some students (22%) tested examples after generating a proof. However, this can only be interpreted with caution. While they may have been checking their proof by using examples, the interview protocol focused on eliciting examples from participants, and thus, these latter examples developed as part of a conversation between the interviewer and student and cannot be considered to be sufficient evidence that the students were not convinced by the generality of their proof.

The relative ordering of the empirical and the valid approaches was affected by a student's initial belief in the truth of the mathematical conjecture. Proofs only occurred before the first example (4 cases of this ordering) for students who believed the conjecture to be true. The remaining students with proofs (n = 4 believed true, n = 10 other reactions) all tested examples before arriving at their proofs. Overall, 78% of the valid proofs were preceded by examples. One student verbalized the approach of using empirical strategies to support proof generation by explaining that his arithmetic with the examples led to the development of his proof. Students were equally likely to produce examples before a proof and after a proof.

A Focus on the Empirical Strategies

The implementation and complexity of the empirical approaches varied across students. The interview assessed ways in which students varied the quantity, diversity (i.e., parity and magnitude), and self-reported typicality of the numbers they tested. As will be seen, the overall complexity of an empirical approach was influenced by a student's initial reaction and whether or not the student generated a valid proof.

Quantity of Examples Tested Overall, students using an empirical approach recognized that they needed to test multiple examples:

> "And the more times you try it, the more likely your study is gonna be right. Or you – but the better answers you're gonna get. So if you tried it with a thousand numbers, you're gonna have better data than if you just tried it with three" (student in 8th grade math).

There were nuances to this approach, however. Students tested fewer examples when the problem was initially believed to be true, particularly on the whole number conjecture (see Figure 3). In addition, students who produced at least one proof tested fewer examples on average across the two conjectures (M = 3.12) than students who produced no proofs (M = 3.86), however this was not

[2] 'Case' refers to a students' response to one of the two problems as the unit of analysis.

[3] As some students produced both narrative and visual proofs on a given case, the proportions sum to more than 1.

significant (p = .22) overall. It does suggest, however, that students view examples and proof as mutually supportive, with fewer examples being required when students demonstrate a conceptual understanding of the logic of the conjecture by producing a valid proof.

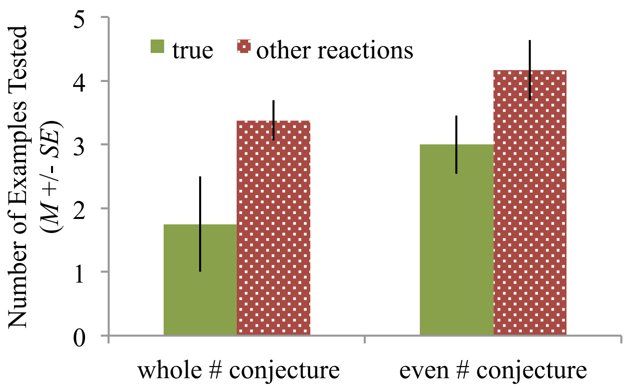

Figure 3: Quantity of examples tested.

Diversity of the Examples Choosing diverse examples allows a stronger test of a conjecture. A student's variations of the magnitude and/or parity of their examples were selected as an objective measure of the diversity of the set of examples. Variation in magnitude was operationalized as having numbers both below and above 20. The domains specified by the conjectures influenced the features that students varied in order to establish diverse sets. On the whole number problem, 74% of the students varied parity while only 37% varied magnitude. However, on the even number problem, 75% of the students varied magnitude.

Students who believed the even number conjecture to be true were less likely to vary magnitude (50%) than students with other reactions (92% varied magnitude). Although this pattern is not repeated for variation in magnitude when testing the whole number conjecture, it does appear in the students' probability of varying parity. Students who believed the whole number conjecture to be true were less likely to vary parity (33% of students) than those with other initial reactions (81%). Thus, believing the conjecture to be true reduced the amount of variation implemented in whichever means the student had selected to vary the example set. The rarity of co-varying both parity and magnitude on the whole number problem set (see Table 1) likely attenuated the effect on this dimension.

Table 1: Covarying magnitude and parity was rare (n).

	Varied Magnitude	No Variation in Magnitude
Varied Parity	4	10
No Variation in Parity	3	2

Thus, it appears that students were less likely to select a diverse set of examples (as measured by parity and differences in magnitude) if they initially believed the problem to be true.

Returning to the subgroup of students with valid proofs, there was less variation in their example sets than for students who did not develop proofs. Students who had proofs were less likely to vary parity on the whole number problem (41% versus 29%) and were less likely to vary magnitude on the even number problem. This represents a similar pattern of findings to the subgroup that initially believed the responses to be true. While the two findings do appear to exist independent of each other, it is also important to remember that proofs were more likely among participants who believed the conjectures to be true. Thus it is the students who were more skeptical of the conjecture who selected more diverse examples. These students seemed to appreciate that a more diverse set of examples provides stronger evidence for a conjecture's truth if a more deductive approach was not available.

Varying Typicality of Examples The examples students generated were coded in terms of the students' self-reported judgments of typicality as well as in terms of the numbers' mathematical typicality. When students explained what made a number typical, they referenced parity, primes, multiples, and magnitude (odd numbers, primes, large numbers, and numbers uncommon in everyday life were considered as unusual). Mathematical typicality was defined a priori by the researchers based on whether properties of the number made it mathematically special within the context of middle school mathematics. Thus, numbers that have identity relations (0,1), are powers of 2, are prime, or are multiples of 5 or 10 were defined as mathematically special. The remaining numbers were coded as mathematically ordinary.

Overall, there was a positive relationship between varying self-reported typicality and varying the coded mathematical typicality, (r (36) = .30, p = .06). Over 70% of the generated example sets had both mathematically ordinary and mathematically special numbers. Of these sets, over half were cases in which the student had reported using typical and unusual numbers; the remaining students (44%) reported using only typical numbers. Among the sets that did not vary mathematical typicality, 8 sets used only mathematically special numbers and 2 sets used only mathematically ordinary numbers. Thus, even though they varied in their self-reported typicality, the tendency was to use unusual numbers even if the mathematical typicality of the example set did not vary.

Thus, it appears that varying the mathematical typicality of numbers was one way students generated sets of examples to test. Using this as a strategy could indicate that underlying conceptual knowledge about the properties of numbers influenced students' example choices.

Was Varying Self-Reported Typicality a Strategy? As the interview formats had some variation, responses to whether varying typicality was intentional and/or a good strategy are collapsed across the two conjectures and reported below at the student level.

All students reported that they tested typical numbers at least once during the interview and that using typical numbers was a good strategy. The majority (75%) of students reported testing an unusual number at least once, and all who were asked (n = 13) indicated at some point that testing unusual numbers was a good strategy.[4] For example, a student taking geometry reported, "if he didn't use unusual numbers, you know, you can never be sure if his property is correct". Interestingly, students who produced a valid proof at least once were more likely to report intentionally selecting these categories of numbers (see Table 2).

Table 2: Proportion of subjects using a self-reported category of numbers who did so intentionally.

	Typical	Unusual
Produced a Valid Proof at least once	0.82	0.88
No Valid Proofs	0.40	0.25

Most (86%) of subjects who reported using both types of examples and were asked about the benefits of using both indicated that using some typical and some unusual was a good strategy. Thus, although students seemed to recognize that varying typicality was an important approach, they did not necessarily do so intentionally, especially if they did not also construct proofs. Such distinctions between using a strategy, using it intentionally, and recognizing it as a good strategy reflect the developing nature of empirical justification approaches within this sample of students.

How would you show someone else?

When students were asked how they would show someone else that the conjecture was always true, they were being asked to implicitly evaluate how convincing their approaches were. This offers insight into how the students value the empirical inductive strategy and the deductive strategies. Focusing on the even number conjecture[5], 11 students did not generate valid proofs. Over half (7) reported that they would use different examples with or without the current examples,

> "I'd try some other examples. Because if I used some ones that you people wouldn't normally use, besides 10, and if I did a little more maybe it wouldn't be or maybe it'd still be true" (student in 6th grade math).

The remaining 4 students without valid proofs reported that they would use the current examples they had generated ("these examples because I don't really get the logic behind it," student in 7th grade math). Thus, the majority of these students did not believe the examples they tested were sufficient to convince someone else of the conjecture's truth. Their empirical approach, while it was sufficient to convince them during the interview, was simultaneously deemed insufficient – necessitating either more logical approaches or further diversification of the examples.

Valid proofs were generated by eight students. Six of them reported that they would use their proof to demonstrate the mathematical property to someone else, although one indicated that examples should precede the proof. The students' explanations of why they would use the proof to show someone else were very clear:

> "It's way more convincing than all that stuff [trying examples]. Now that I can like see how it works out instead of just like finding, oh, it does work out." (student in 8th grade math)
>
> "I think I find the second one I said more convincing 'cause it's a little bit more in general. And it's not using like one specific number. It's giving a rule kind of. [It's] using a variable to some extent." (student in 7th grade math)

When students had both the inductive and deductive approaches available to them, they found the deductive proofs more convincing. They understood the value of a valid proof in justifying a conjecture's general truth. Thus, although students of this age group are known to rely on empirical methods and did so in this study, they also showed a developing understanding of the benefits of proof.

General Discussion

Empirically-based inductive strategies to justify mathematical conjectures can co-exist and complement the more formal deductive strategies. Strategic use of examples may be important at the beginning stages of mathematical justification. While over half of the interviewed students did treat empirical approaches as if they were valid, they used the empirical approaches in a strategic manner by varying the quantity, diversity (parity and magnitude), and typicality of the tested numbers. Despite the limited sample size, the analyses of the students' thinking processes during the interviews revealed rich and strategic approaches to justifying mathematical conjectures. Given the preponderance of empirical-based reasoning demonstrated here and in other studies, such an in-depth examination of the use of examples is critical for understanding students' current approaches and developing ways to leverage these to support deductive reasoning.

Looking across the different measures, students who generated a convincing, deductive generalization tested

[4] Three of the five students who reported only testing typical numbers were asked about the strategy; they thought testing unusual numbers would not be a good strategy. Their explanations centered on the computational ease of the typical numbers.

[5] Only four students were asked on the whole number conjecture. They had all generated proofs and indicated that they would use the proof to show someone else.

fewer examples, intentionally selected the typicality of the numbers they did test, and had less diversity in their example set. The students generating proofs correctly placed less importance on strategic use of empirical strategies than those who only used an empirical approach. However, it is not a 'problem' that the students who generated proofs had less variation and fewer examples. In fact, their reports of more frequent intentional example choices suggest that they picked examples strategically and they recognized there was no need for further empirical tests.

The students who did not generate proofs also reasoned strategically. They generally used multiple examples and valued diversity in their example sets. At the same time, however, other aspects of justification are still developing. For example, students rarely varied parity and magnitude within the same problem, despite believing variation to be good. Further, while they considered their chosen examples to be varied, students who did not produce a proof often said they would use even more diverse numbers to demonstrate the truth of the conjecture to someone else. The use of diversity as a cue for inductive generalization is developing during the elementary school years in biological reasoning (Rhodes et al., 2008); perhaps a similar transition occurs in math.

The exploratory nature of the interviews and the limited number of conjectures and proofs prevent full consideration of how empirical approaches interact with proof generation. Even if one is to take the perspective that inductive strategies reflect a shortcoming in the long-run, understanding what students are actually doing in the short-term could aid the development of their mathematical knowledge. Empirical justifications can reflect important mathematical reasoning in their own right. Recognizing that some sets of examples are better than others might lead towards considering whether other types of approaches are better than empirical approaches. Strategic use of examples could develop from the recognition of the weaknesses that exist when testing a limited number of similar examples. This recognition can be harnessed to suggest that similar weaknesses also exist even when you strategically choose particular examples, thus supporting the move to deductive strategies. Our next steps with this research include surveying students in order to more fully understand the role of typicality and diversity in their choice of examples.

In sum, using a combination of empirical and deductive strategies, almost all students in the study were correctly determined that the conjectures were true. The analysis presented here revealed that students could use examples to attempt to falsify conjectures, demonstrate that conjectures work, and perhaps identify patterns and develop a more general proof. Further, many students strategically chose their examples to test, suggesting that they were thinking critically about the underlying properties of the number system and the ways in which typical, unusual, and diverse examples can be used to support mathematical inference-making. As strategic use of examples is an under-researched area, the concepts that emerged during these interviews are important avenues for future research.

Acknowledgments

The research is supported in part by the National Science Foundation under grant DRL-0814710. Special thanks to the students who participated in the interviews.

References

Alcock, L. J. (2004). Uses of example objects in proving. In M. J. Hines & A. B. Fuglestad (Eds.), *Proceedings of the 28th annual conference of the International Group for the Psychology of Mathematics Education (Vol. 2).* Bergen, Norway: Bergen University College.

Gelman, S. A., & Kalish, C. W. (2006). Conceptual development. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol 2, cognition, perception, and language (6th ed.).* Hoboken, NJ: John Wiley & Sons Inc.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological review*, *111*, 3-32.

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

Kloosterman, P., & Lester, F. (2004). *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Koedinger, K. R. (1998). Conjecturing and argumentation in high-school geometry students. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space. Mahwah, NJ: Erlbaum.*

Knuth, E., Choppin, J., & Bieda, K. (2009). Middle school students' production of mathematical justifications. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective*. New York, NY: Routledge.

Knuth, E., Kalish, C., Ellis, A., Williams, C., & Felton, M. (In press). Adolescent reasoning in mathematical and non-mathematical domains: Exploring the paradox. To appear in V. Reyna, S. Chapman, M. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making*. Washington, DC: American Psychological Association.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Porteous, K. (1990). What do children really believe? *Educational Studies in Mathematics, 21,* 589-598.

Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108*, 543 – 556.