Educ Stud Math (2008) 68:247–261
DOI 10.1007/s10649-008-9114-8

Learning opportunities from group discussions: warrants become the objects of debate

Keith Weber · Carolyn Maher · Arthur Powell · Hollylynne Stohl Lee

Published online: 15 April 2008

Abstract In the mathematics education literature, there is currently a debate about the mechanisms by which group discussion can contribute to mathematical learning and under what conditions this learning is likely to occur. In this paper, we contribute to this debate by illustrating three learning opportunities that group discussions can create. In analyzing a videotaped episode of eight middle school students discussing a statistical problem, we observed that these students frequently challenged the arguments that their colleagues presented. These challenges invited students to be explicit about what mathematical principles, or warrants, they were implicitly using as a basis for their mathematical claims, in some cases recognize the modes of reasoning they were using were invalid and reject these modes of reasoning, and in other cases, attempt to provide deductive support to justify why their modes of reasoning were appropriate. We then describe what social and environmental conditions allowed the discussion analyzed in this paper to occur.

Keywords Discussion · Discourse · Mathematics education · Probability · Statistics

1 Introduction

A central issue in mathematics education is the role that discussion and discourse should play in mathematics classrooms (Sfard 2001). As discussions "allow students to test ideas, to hear and incorporate the ideas of others, to consolidate their thinking by putting their ideas into words, and hence, to build a deeper understanding of key concepts" (McCrone 2005, p. 111), many researchers and influential organizations recommend that discussion play a prominent role in reform-oriented mathematics classrooms (e.g., Balacheff 1991; Bauersfeld 1995; NCTM 2000). However, simply having students discuss mathematical ideas does not

K. Weber (✉) · C. Maher · A. Powell
Graduate School of Education, Rutgers University, New Brunswick, NJ 08840, USA
e-mail: khweber@rci.rutgers.edu

H. S. Lee
North Carolina State University, Raleigh, NC, USA

guarantee that meaningful learning will occur. In analyzing the value of a mathematical discussion, one must consider, among other factors, the nature of the students' interaction (McCrone 2005) and the depth and quality of the mathematical ideas being discussed (Hiebert and Wearne 1993). Exactly how discussion should be promoted in mathematics classrooms or what types of discussion are likely to provide learning opportunities are important open questions (Lampert, Rittenhouse, and Crumbaugh 1998; McCrone 2005).

One goal of this paper is to contribute to this debate by illustrating specific ways in which discussions can contribute to mathematical learning. In the mathematics education literature, several such accounts have been proposed. Discussion can objectify students' previous experiences, thereby making these experiences the subject of analysis (Cobb, Boufi, McClain and Whitenack 1997), encourage students to take a more reflective stance on their mathematical reasoning (Manhouchehri and Enderson 1999), require students to consolidate their thinking by verbalizing their thoughts (McCrone 2005), and help students learn to communicate mathematically and participate in a wider range of mathematical argumentation (Lampert and Cobb 2003; Sfard 2001). We contend that group discussion can also facilitate learning by inviting students to be explicit both about the ways in which they make new claims from previously established facts and about the standards they are using in deciding whether an argument is acceptable. Challenges from classmates can encourage students to debate whether a particular method of argumentation is appropriate and provide students with the opportunity either to justify their methods when their reasoning is sound or revise or abandon their methods when their reasoning is flawed. A second goal of this paper is to describe social and environmental conditions that invite productive discussions of this type.

2 Theoretical perspective

2.1 Toulmin's model of argumentation

Krummheuer (1995) and others (e.g., Yackel 2001; Rasmussen and Stephan 2008) have argued that Toulmin's (1969) model of argumentation can be a useful analytical tool for understanding the progression of students' arguments during collective discourse. In Toulmin's model, an argumentation consists of three essential parts, called the core of the argument: the claim, the data, and the warrant. When an individual presents an argument to his or her community, he or she is trying to convince the audience of a particular *claim*. To support the claim that is being made, the individual typically presents evidence, or *data*. The audience may ask the individual presenting the argument to explain why one should deduce the claim being made from the data being presented. In Toulmin's scheme, such an explanation is referred to as a *warrant*. It is possible that an audience will accept the data presented in an argument but not be convinced that the conclusion of the argument necessarily follows from the data. In other words, the audience is questioning the validity of the warrant. When this occurs, the one presenting the argument may present support, or *backing*, for the warrant (cf., Stephan and Rasmussen 2002). Toulmin's model of argumentation consists of two additional components – a *modal qualifier* qualifies the conclusion by expressing degrees of confidence and a *rebuttal* states conditions under which the conclusion does not hold. Inglis, Mejia-Ramos, and Simpson (2007) argue that although these components generally are not used by mathematics educators when analyzing argumentation, considering these qualifications can provide a more comprehensive description of individuals' argumentation and reasoning processes.

2.2 Uses of Toulmin's model in mathematics education research

Toulmin's model has been used in mathematics education primarily in one of two ways. First, some researchers use Toulmin's model as a tool to analyze students' evolving conceptions by documenting their collective argumentation (Krummheuer 1995; Yackel 2001; Stephan and Rasmussen 2002). For instance, Stephan and Rasmussen (2002) treat an idea as "taken as shared" by a group of students if it is used as data in an argumentation and no student in the group challenges this data. Second, other researchers use this model to categorize or assess the quality of a specific mathematical argument (Pedemonte 2007; Weber and Alcock 2005; Inglis et al. 2007)–sometimes as a normative tool with which to judge whether an argument constitutes a mathematical proof (Pedemonte 2007; Weber and Alcock 2005) or to describe the progression of an informal argument to a valid deduction (Pedemonte 2007).

In this paper, we will use Toulmin's scheme in both of the senses described above. We use Toulmin's scheme in the first sense to illustrate how challenges to students' arguments can lead the students to be explicit about the warrants they are using and to have the class as a whole collectively debate whether, and under what conditions, the warrants are appropriate. Often, when a student presents a mathematical argument, he or she is not clear about what warrants are being used. Indeed, that student may be employing warrants implicitly without ever having considered whether these warrants are valid (Rasmussen and Stephan 2008). Challenges to these arguments from the student's classmates invite the student to be explicit about the warrant being employed and provide backing to support the warrant's legitimacy. There may also be cases in which students obtain conviction in a warrant via empirical or abductive reasoning [Here, abductive reasoning refers to instances in which students infer a mathematical explanation from observing a small number of situations. As Johnson-Laird (2006) notes, "when our reasoning goes beyond the premises to increase the amount of information, we are making an induction […] some theorists refer to inductions of explanations as abductions" (p. 175)]. If a student presents an argument that uses such a warrant, a challenge from a classmate may invite that student (or the class as a whole) to construct a deductive argument as backing for that warrant (i.e., a logical justification for why the students' inductive inference must necessarily hold). Hence, discussions can lead students to refine their arguments so they are mathematically tighter by constructing deductive backings for their warrants (cf., Pedemonte 2007). Finally, there may be cases in which students are basing their arguments on warrants that are not generally valid. In these cases, challenges to students' warrants can lead the class to recognize the that the warrant being used is not legitimate and to abandon that mode of reasoning. In all cases, challenges to students' warrants can shift students' reasoning away from the particular situation they are investigating and facilitate a move toward more general and abstract mathematical principles.

In this paper, we present a discussion in which students debate whether various six-sided dice are fair based on data they obtained from running computer simulations. The claims students made were judgments on the fairness of each of the dice used in our study. The data (in Toulmin's sense) usually, but not always, consisted of data (in a statistical sense) obtained from running simulations in which the die under consideration was rolled multiple times. What was most interesting to us as researchers was neither whether a student believed a particular die was fair (the student's claim) nor whether the output of the computer simulations that the student used to justify this conclusion (the data) but rather what principles the student was using to draw conclusions from examining data (the warrant) and the ensuing debates about whether these principles were legitimate (the debate about warrants and the construction of backings) and under what conditions were these principles appropriate to apply (the issue of potential rebuttals and appropriate qualifications to the warrant).

3 Research context

3.1 Research setting

The research reported in this paper occurred in the context of the "Informal Mathematical Learning" research project[1]. In this project, an innovative after school program was implemented at Hubbard Middle School in Plainfield, New Jersey, an economically depressed urban area. Ninety-eight percent of the students who attend Hubbard Middle School are African American or Latino. Twenty-four sixth grade students, all African American or Latino, volunteered to participate in the Informal Mathematical Learning program.

This project was a longitudinal study, spanning three years. Students' participation in this study began as they entered sixth grade and ended when they completed eighth grade. During their participation in the program, students met after school with the researchers of the program for 60 to 90 minutes. There were approximately 20 such meetings during each academic year. A subset of the students also participated in sessions in the summers following the first and second years of the study. These summer sessions consisted of four three-hour meetings over the course of one week. The data in this paper came from the last meeting of the second summer session. Eight students participated in this summer session, and these eight students were representative of the 24 students who participated in our study.

3.2 Goals of the research study

The goal of our research study was to understand how students constructed mathematical ideas while completing interesting mathematical tasks when certain social and environmental conditions were in place. Therefore, the goals of our study differ from other teaching experiments where researchers have a particular concept that they want the students to construct, often by traversing through an anticipated learning trajectory (e.g., Simon 1995). We did not enter this research with specific ideas that we expected students to learn, nor did we have particular stages that we wanted students to traverse. Students' ways of reasoning and how they constructed their knowledge were the results of our research, but not our preconceived goals. To emphasize the contrast, we refer to this study as a "learning experiment", not a "teaching experiment". Being in an after-school setting, rather than in a classroom with a fixed curriculum, allowed us to have this focus.

In our study, we attempted to have our students complete mathematical tasks under the following conditions: students were encouraged to work collaboratively; students were expected to justify their mathematical reasoning to their classmates; students were arbiters of whether or not a solution was correct and a solution was only regarded as correct if the students all agreed that "it made sense"; researchers received all student participation positively and did not judge the quality or correctness of students' reasoning; students were

[1] The "Informal Mathematical Learning" Project, led by Carolyn Maher, Arthur Powell, and Keith Weber, is supported by the National Science Foundation ROLE Grant REC0309062. The views expressed in this paper are those of the authors and not necessarily those of the National Science Foundation.

given extended time to work on mathematical tasks; students were given several tasks that concerned the same mathematical ideas, giving students the opportunity to revisit their previous work and refine their mathematical thinking. The researcher's role in this environment is to provide students with interesting and motivating tasks, encourage students to work collaboratively and justify their solutions to their classmates, have students make their reasoning public, and have students carefully attend to the reasoning of others with the aim of making sense of this reasoning. A more complete discussion of these conditions, how they were established, and the researchers' role in our research environment are discussed in Radu, Tozzi, and Weber (2006) and Weber et al. (submitted for publication). In the concluding section of this paper, we describe the researcher's role in the analyzed discussion and analyze how the classroom conditions allowed the productive discussion that we analyzed to occur.

3.3 Probability and statistics tasks using *Probability Explorer*

During the first and second summer sessions, students worked on probability and statistics tasks, often using the *Probability Explorer* computer software (Drier 2000) as a tool to help them complete these tasks. To use Probability Explorer, one initially models a statistical situation. For instance, a researcher may "create" a biased six-sided die that rolls 6's twice as often as it rolls any other number. After the situation is modeled, students can use Probability Explorer to run simulation experiments about the modeled situation. Given the biased die described above, a student could command the software to roll the die 100 times. Probability Explorer would tabulate the results of running 100 trials and, at the student's command, could represent the result in a variety of ways, including as a table (stating the number of times each outcome occurred, as well as the percentage of times each outcome occurred and the ratio between the number of times the outcome occurred and the total number of trials), a pie chart, or a bar chart.

3.4 The Schoolopoly task

The Schoolopoly task (cf., Stohl and Tarr 2002; Tarr, Lee, and Rider 2006) was the last task given to students at the end of the second summer session.

1. Students were told that their class had designed a new board game entitled "Schoolopoly". They needed to choose a dice company who would manufacture the dice to be used for the game. They had six dice companies from which to choose. However, some of these companies had a reputation for making unfair dice. Their task, as a class, was to determine which of the six dice companies made fair dice. Students were split into pairs and each pair of students was assigned two or three dice companies. Each dice company was assigned to exactly two pairs of students. Prior to the experiment, dice from each of the six companies were modeled on each student computer using the Probability Explorer software. The students' initial task was to use Probability Explorer to determine whether or not each dice company they had been assigned made fair dice.
2. For each company that a pair of students was assigned, that pair was asked to create a poster that summarized their findings. The poster was to include their judgment on the dice company (i.e., whether they believed the die they inspected was fair or unfair), the data that they collected that supported their conclusion (they could choose which data to present and could represent the data in the best way that they saw fit), and a written explanation for why they reached the conclusion that they did.
3. Students' posters were arranged so that the two posters that evaluated each dice company were placed adjacent to each other. The students were asked to visit the

posters for each dice company, take notes on these posters, and then choose which dice company from whom they would buy dice.

4. After each student examined each poster and made his or her selection, students were asked to come together to discuss which dice company made the fairest dice. This 30-minute discussion is the subject of the research reported here.

3.5 Analysis

The data were analyzed in a manner consistent with the first stage of the research methodology prescribed by Rasmussen and Stephan (2008). The videotape of the group discussion was viewed repeatedly by members of the research team to get a strong sense of the data. The discussion was then transcribed. The data were parsed into specific student argumentations. Each argument that a student presented was coded according to Toulmin's (1969) scheme; each argument was coded in terms of the claim being made, the data to support the claim, and, when given, the warrant for how the data implied the claim. In cases where no warrant was provided, we sometimes would make a note in our codings with our interpretation of the warrant that the student appeared to be implicitly using. Each coding and all interpretations were discussed within our research team until all disagreements were resolved. We also coded all challenges to a student's argument by what part of the argument – the claim, data, or warrant – was being challenged. In cases where the warrant was challenged, we analyzed how the student was rebutting the presented argument, noting whether this rebuttal was challenging the legitimacy of the warrant itself, or only arguing that the warrant might not be applicable in the situation in which it was being used – that is, the warrant would need to be qualified in such a way that certainty cannot be obtained from its application (cf., Inglis et al. 2007). The result of our analysis was a coded chronological account of the arguments and challenges that the students raised during the 30-min discussion.

4 Results

Below we present a description of the discussion that took place between the students in discussing which die was fair. We organize our description by presenting an analysis of six episodes in which students are engaged in an argument. In our discussion, we often reference the posters to which the students were referring. All posters can be viewed from the following website: http://www.rbdil.org/newposters.

4.1 Episode 1: Chris' consensus-based argument

The first episode occurs at the beginning of the discussion. The researcher begins by asking Chris which company he would buy from:[2]

R1: Chris, we're going start over here with you. What company should we buy from and why?

[2] The transcripts were lightly edited for the sake of brevity and to increase their readability. We omitted short segments of texts, including stutters, off-topic arguing, and text with no mathematical content. Such omissions are denoted by the bracketed ellipsis […]. At no point did we add words or change the words that were spoken.

Chris: Delta's Dice.

R1: Delta's Dice. Why?

Chris: Because, if you look at both of them, they both like, like really explain the same thing. Like to me, I thought the first poster and the second poster were like about the same thing. They really explained it […]

R1: OK. Out of how many trials? What were they doing, do you remember?

Chris: Uh…. I think it was 600. But I don't know.

R1: OK. So it seems to me that one of the reasons why you're picking Delta's Dice is because the two groups agree.

Chris: Uh-huh. Because, like, in other ones, like, one didn't agree and one did agree, or sometimes they didn't really explain enough.

We coded this excerpt in the following way:

Claim: Delta's Dice is fair

Data: Both posters agreed that Delta's Dice was fair and presented similar arguments. Posters for other dice did not agree, or did not give thorough explanations.

Warrant: None provided. We inferred Chris' warrant to be that if both posters for a die agreed, it was reasonable to conclude that die was fair.

Chris seems to be accepting that Delta's Dice was fair based on the authority of the posters that he examined, but not on the data or arguments contained in the posters. When the researcher asked Chris for details about the poster ("Out of how many trials? What were they doing?"), Chris could not recall details of the poster with confidence. He also indicated that the lack of agreement in the posters for other companies was a central reason for why he chose Delta's Dice.

After Chris gave his argument, Chanel challenged it:

Chanel: I just have a real quick question. Why does on the one scribbly and stuff, why does it say that one is lower, one might be lower and the rest are higher, and why, how is that fair? Yeah, I don't get it […] That one, it might be lower, and the rest, the rest is just higher. So, how is, I don't get it, how is that fair?

In this excerpt, Chanel is questioning whether the data in one of the posters for Delta's Dice really is evidence that Delta's Dice is fair. Several other students raised questions similar to Chanel's. Chanel's challenge is more sophisticated than Chris' argument since she examines the data and argument presented in the poster and not just the conclusion expressed in the poster. However, Chanel's challenge was to the conclusion that Chris inferred (that Delta's Dice was fair), but not to the reasoning he used to reach his conclusion. She did not question whether it was appropriate to conclude a die was fair because both posters in support of it claimed it was; rather she challenged Chris' claim because she reached a different conclusion. Next, Danielle and Keisha then presented similar arguments that Delta's Dice was fair, primarily relying on the fact that the two posters evaluating Delta's Dice both concluded that it was fair.

4.2 Episode 2: Chanel's challenge to Tiffany's uniform distribution argument

Tiffany deviated from the previous students by arguing that Calibrated Cubes would be the die most likely to be fair:

R1: OK. Tiffany, what do you think?

Tiffany: I picked Calibrated Cubes.

R1: OK. Can you tell us why?

Tiffany: Because, I think it's fair because all the numbers were even, 'cause when I looked at the charts, all the numbers had 11, I think.

Here, Tiffany is referring to a table on one of the posters that showed that on one of the trials, 1, 2, and 3 each occurred exactly 11 times. The table on the poster was arranged in such a way that the number of times a 4, 5, and 6 occurred were not shown. We coded this argumentation as follows:

Claim: Calibrated Cubes is a fair die.

Data: The table for Calibrated Cubes shows that 1, 2, and 3 each occurred exactly 11 times.

Warrant: None given. We inferred that Tiffany was using a uniform distribution argument–if, in a given trial, the die produced the same output for each of the numbers, that die was (likely) fair. (Tiffany's repeated use of the phrase, "I think", suggests to us that she may be qualifying her argument).

Tiffany differed from the previous students in that the data that she mentioned in support for her claim did not concern the judgment of the posters that she inspected, but rather their contents. Chanel immediately challenged Tiffany's argument:

Chanel: I think that like on Calibrated Cubes it just showed three 11's. It didn't show all of the cubes. 'Cause there were three more cubes, and those could have been 12, 13, or 14, or any other number. And it didn't show all the numbers, it showed the three 11's. How do we know it wasn't like 34 or something on the other ones? […]

Tiffany: But I have a question. Whoever that is, what was the other numbers? You don't have to lie.

In this excerpt, Chanel is challenging Tiffany's implicit warrant by proposing a potential rebuttal. Tiffany's argument would only hold true if the data showed that all of the numbers occurred equally often, but the data Tiffany cited only showed that 1, 2, and 3 occurred equally often. Chanel noted that one of the missing numbers could conceivably have been as large as 34; if so, Calibrated Cubes would have made unfair dice. Tiffany's request to know the missing values of the table suggests that she is attending to and appreciates the merits of Chanel's challenge. Finally, we note Chanel's challenge here differs from the one that she and others posed to Chris. Chanel did not simply challenge Tiffany's claim because she arrived at a different judgment; here she is challenging the validity of Tiffany's reasoning.

4.3 Episode 3: Jerel's counterargument stressing the importance of sample size

Jerel challenges Tiffany's argument for a different reason than Chanel.

Jerel: Well look. They only ran it 80 times. You'll never know if another number is gonna come up and pass it. Even though it was even, they ran it a small amount of times. You need to run it a lot of times. […] just like when we were doing Delta's Dice, we had ran it, I think a hundred times, and one number won by a lot. But when we ran it like one thousand times and all that, other numbers won […] The reason they got that all is because they had a little bit amount of numbers that they ran it… I guarantee you if you ran it like 500 times, it would have been different. You ought to say it was unfair.

In this exchange, Jerel is proposing a different rebuttal to Tiffany's argument. Tiffany appears to basing her reasoning on what Tversky and Kahneman (1983) refer to as *the law of small numbers*, the commonly held yet erroneous belief that the properties of an unknown distribution can be assessed with a relatively small number of observations. In his challenge, Jerel is explicitly stating and then questioning this assumption. Citing his experience in seeing discrepancies in his data when he ran simulations with 100 and 1000 trials, Jerel puts forth a counterclaim that is related to the *law of large numbers* – informally that increasing the number of trials will tend to make one's empirical approximation of an unknown distribution more accurate. Jerel's counterclaim that one needs to run a large number of trials to obtain a reliable estimate goes beyond the specificities of Tiffany's argument, focusing on a broader statistical issue.

4.4 Episode 4. Chris, Jerel, and Terrill debate the importance of sample size

For the next six minutes, the students debated whether Jerel's claim – that one needed to run many trials to reach legitimate conclusions – was valid. One excerpt from this debate is given below:

Chris: I said, well, that poster said that if you want your money back and anything you should run it more. You're not going to roll only 5 times and stop playing. So, if you're playing for your money and anything, you're going to want to win your money back or win more money. That's why. So you never know, you never know. […]

[Later in the discussion, Jerel was asked how many times he and Chris ran their simulations]

Jerel: One thousand times. But would you have, all right, would you have start playing a game which I have rolled a hundred times and say, 'Oh I'm boring [sic], I want to eat, I don't want to play no more' would you do that? Or you would have kept going, right?

Terrill: I don't understand how gambling comes into this because I did a lower number but that, with a lower number it shows you which ones that, like, with a lower number it shows you which ones like, OK they did it a thousand times but do you seriously think that when they're playing a game they're going to roll a thousand times? No. Well, 80 is a more reasonable number, that's why I chose 80.

Chris tries to support Jerel's argument with a gambling metaphor. As data, he claims that if one was gambling – a context in which die rolling is frequently employed – you wouldn't stop at 80 trials, but would continue rolling. Chris does not provide an explicit warrant for how one's gambling decisions relate to the sample sizes one needs to draw reliable statistical conclusions, but he appears to be saying that there should be a connection between the way that one rolls dice in real life contexts and how one rolls dice to make statistical judgments. In the same excerpt, Jerel makes a claim similar to Chris – if you are playing a game, you are not likely to stop after only 100 rolls.

Terrill challenges both the (hypothetical) data that Chris presented – "Do you seriously think that when they're playing a game they're going to roll a thousand times? No. Well, 80 is a more reasonable number" – and Chris' implicit warrant – "I don't understand how gambling comes into this".

4.5 Episode 5: Chanel's argument against the importance of sample size

Several other counterarguments to Jerel's claim were given by other students. One of the most interesting counterarguments was given by Chanel:

Chanel: No what I'm saying, um, like, Jerel and Chris always run them for a thousand or something out of a thousand. How do you know you can't run them out of a thousand-five hundred? How do you know that's not high enough? You're always going to a thousand. I've never seen…

[Chanel stops in mid-sentence. Students talking on top of each other]

Someone: 100 is a high number.

Terrill: That don't automatically mean you're right, just because you use a higher number.

Chanel: Let me finish. They always run it up to a thousand. How do you know you can't run it up to 3000, or um, it has to go higher because, just because you go up to a thousand, that's the only number I always see them do. How come you can't run it to 50, and you can't find the answer then? Or write it from 10, or whatever, and find the answer then?

In carrying out earlier simulations, Chris and Jerel would regularly run one thousand trials, on the assumption that a large number of trials will yield a more reliable conclusion (see episodes 3 and 4). However, based on the assumption that "more is better", Chanel argues that Chris and Jerel should be running more than 1,000 trials, perhaps 3,000, since that will give a more reliable estimate still. If one does not need to run 3,000 trials and could still obtain a reasonable estimate with 1,000 trials, why can't one do the same from a smaller number of trials, such as 50 or even 10?

4.6 Episode 6: Terrill's counterargument to Kianja's consensus-based argument

Later in the interview, the researcher was asking some students why they did not pick certain dice as fair. Kianja gave the following explanation:

R1: Um, those of you, I mean all of us went and visited Dice R Us. Why didn't you pick it? Kianja?

Kianja: Well, for one I didn't pick it because it was 2 different opinions, and if you have 2 different opinions it's hard to pick like which one is right or whatever. Like some might say it was unfair and some might say it was fair.

We coded Kianja's argumentation as follows:

Claim and qualifier: We cannot have confidence that Delta's Dice is fair.

Data: One poster said Dice R Us was fair and one said that it was not fair.

Warrant: If two posters express different opinions, it is difficult to judge which opinion is correct.

Kianja's argument is similar to the argument Chris presented in the beginning of the session (see episode 1) in that both students refer to the conclusions of the posters, rather than to the data that the posters presented or the arguments that the posters made. Terrill then challenges the validity of such a warrant.

Terrill: Can I say something here? […] OK, now, you all are picking it based on two people saying that it was right?

Kianja: No, I just said…there's two different arguments, and…'cause it was two different arguments, and both of their arguments were convincing, so how am I going to know which one to pick? Because the fair one was convincing me that it was fair, and the unfair one was convincing me that it was unfair, and I didn't know which one it was […]

Terrill: Exactly. That's what I was saying […] OK, just because two people said it was fair doesn't mean it's fair, so why would we pick the one like, OK, say if two people was wrong, what if those two people was wrong, then what? You was wrong. Just because two people said it was fair don't mean you got to pick it.

Kianja: But would you want an unfair dice?

Terrill: How do you know it was unfair? Do you know, are you positive that it was right?

In this excerpt, Terrill challenges Kianja's warrant by questioning whether one can draw conclusions solely by looking at the opinions offered in the posters. Terrill presents a possible rebuttal to Kianja's argument–the posters in question could present an ungrounded or false argument ("What if those two people was wrong, then what? You was wrong"). It is interesting that Terrill went beyond the specific context in which Kianja reasoned – claiming a decision could not be reached because the two groups disagreed – to the more general issue of basing conclusion on whether two posters agreed or disagreed. After this episode, no student made an argument for the fairness of the dice based on the agreement of the posters.

4.7 Follow-up intervention

The discussion concluded without the group reaching a consensus on whether a data set with a large number of trials is needed to draw reliable conclusions. At this point, the researcher attempted to capitalize on this tension by suggesting the students run an experiment with Calibrated Cubes – a company that several students believed made fair dice. In this intervention, the researcher's role was more proactive than at most other points in our study. Together, the researcher and the students co-designed an experiment with Calibrated Cubes in which a large number of trials were run. Periodically, the researcher would pause the experiment so the students could take stock of the trial distribution as the experiment progressed. During the experiment, the proportion of times that each number occurred varied widely at first, but as the number of trials grew larger, the proportions appeared to level out. The numbers of 2, 4, and 6's rolled were roughly equal. So were the numbers of 1's, 3's, and 5's. However, each even number was rolled about 50% more often than each odd number. The pie chart and bar chart displayed while the experiment was running provided the students with the opportunity to observe this. Prior to the experiment, Terrill and Danielle believed Calibrated Cubes was fair. After the experiment was completed, the researcher asked these students if they had changed their minds.

R1: Now, a couple people said that Calibrated Cubes is the company to choose […]

Jerel: And now I know they changed their minds.

Chanel: Danielle, you changed your mind, right?

Terill: I changed my mind. I like Dice R Us.

Danielle: [singing] I changed my mind.

R1: What did you change it to, Danielle?

Danielle: I don't want Collaborated Cubes no more […]

Chanel: Why you don't like it no more?

Danielle: 'Cause now I can see, that…

[students talking over one another]

Danielle: Now I can see what, uh, basically what, um,

Jerel: We were trying to say!

Danielle: Yeah, what they were trying to say, 'cause it, it was a low number, but if you did try the high numbers more than you try what we did, um, you know you will come up with a different answer. So that was like a one-time probably lucky thing.

4.8 Summary

The first three students who offered their decision on which die was fair chose Delta's Dice, primarily because the two posters that inspected Delta's Dice both found Delta's Dice to be fair and no other dice company shared a similar level of agreement. The students making these arguments discussed the content of the posters only briefly, if at all. Only one of the three arguments was challenged. The challenges posed were not based on the reasoning that the student used, but on the conclusion that he reached. Tiffany was the first student to present an argument based on the data presented in the posters that she inspected. Challenges to Tiffany's arguments were based on potential rebuttals to the warrant that she appeared to be using to draw conclusions. A central challenge to Tiffany's data was whether it was appropriate to draw conclusions from a simulation that used a relatively small sample size. The issue of sample size led to a lengthy and lively debate. Subsequent challenges to students' arguments were based on the warrants that the students were using, or appeared to be using implicitly. In particular, arguments based on the conclusions of the posters but not the data presented in the posters were challenged by other students, such as Terrill's challenge to Kianja presented in Episode 6. Although a consensus about the importance of sample size was not reached during the group discussion, the researcher's follow-up intervention convinced at least one student, Danielle, that a large number of trials was needed to draw reliable statistical conclusions.

5 Discussion

5.1 Learning in mathematical discussions by debating warrants

The discussion presented in this paper was ostensibly about answering a specific statistical question–which dice company produced the fairest dice. However, as the discussion progressed, the focus shifted to what types of reasoning were appropriate when deducing general claims from empirical evidence. This illustrates how group discussion can provide students with learning opportunities by making students' implicit warrants explicit and the objects of debate. By their nature, warrants are more general and abstract than the specific claim being argued. When warrants become explicit and the subject of debate, their status changes. The warrants become the claim to be justified, engaging students in a higher level of mathematical reasoning.

In this section, we will document specific instances in which we believe learning occurred and then discuss how the group discussions provided the opportunity for this learning. First, in Episode 2, Chanel's challenge to Tiffany called attention to the fact that Tiffany was concluding that a die was fair because three of the six numbers occurred equally often. Similarly, Jerel's challenge to the same argument in Episode 3 highlighted that Tiffany was implicitly using the law of small numbers (Tversky and Kahneman 1983). These challenges made the implicit warrants used in Tiffany's argument public, which provided Tiffany and her classmates a better understanding of the argument that Tiffany was proposing. Second, Chanel's challenge in Episode 2 led Tiffany to recognize a limitation in her argument. After Chanel's challenge, Tiffany asked the author of the poster for the frequency of the missing numbers, which we took as evidence that Tiffany realized she would need this information to form a reliable conclusion about the fairness of the die. Similarly, several students contended that if two posters reached the same conclusion, then the conclusion of the poster was likely correct; otherwise the reliability of both posters was in doubt. Terrill's challenge in Episode 6 allowed students to hear an argument for why this reasoning was inappropriate. After Terrill's challenge, no student offered a counterchallenge or made a subsequent argument based on poster agreement. We took this as evidence that students recognized the limitations of arguing based on the conclusions, rather than the content, of the posters, at least within this specific context. These instances illustrate how challenges to students' argument can lead students to recognize that a particular mode of reasoning is illegitimate.

Third, Jerel's challenge to Tiffany's reasoning in Episode 3 made the notion of sample size, a central concept in statistics, the object of discussion. From Jerel's comments in episode 3, we inferred that Jerel and Chris had previously observed that one could not be sure which number would occur most often after 100 trials but the results were more reliable if 500 trials had been run. To account for this, Jerel and Chris hypothesized that more trials yielded more reliable results. This appears to be an instance of abductive reasoning; Chris and Jerel inferred a general mathematical principle – that running more trials yields more reliable results – to account for their observations that the outcomes of running a small number and large number of trials sometimes produced contradictory results. As Pedemonte (2007) notes, abductive reasoning is appropriate for the generation of conjectures and informal argumentation, but it is often desirable to treat this form of reasoning as a first step in forming a deductive argument that serves as stronger mathematical support for the claims being debated. Challenges to Jerel invited students to produce such a deductive argument in support of or against the importance of sample size; attempts at such deductive argumentation are found in episodes four and five. This illustrates how considering the warrants used in students' arguments provides students with the opportunity to debate more general mathematical principles.

5.2 Creating environments that encourage productive discussion

Creating environments where discussions can provide opportunities for mathematical learning is a pedagogical issue in need of research (Lampert et al. 1998; McCrone 2005). As documented in this paper, we believe that a productive discussion occurred in our study – the students were all active in the discussion, presenting and defending mathematical and statistical claims, carefully attending to each other's arguments, and challenging these arguments if they did not find them convincing – and this discussion led to substantial learning opportunities. Of course, we do not claim that all discussions would lead to this outcome, and we conclude this paper with a necessarily speculative analysis on what actions by the teacher–researcher and what social and sociomathematical norms supported the discussion that we analyzed.

As emphasized earlier in this paper, we did not enter this study with strong notions of what we wanted students to learn or the mechanisms by which their learning would occur. Rather, we attempted to put desirable conditions into place and see what learning patterns emerged. From the beginning of our study, all student contributions were received positively by the researchers. The researchers never evaluated the mathematical correctness or sophistication of the arguments that students prevented. Lubienski (2000) notes that some students, particularly those with low-SES status, believe the purpose of discussion is to allow teachers to judge the students' mathematical knowledge. Students with such a belief think their role in discussions is to present the mathematics that they know and conceal what they do not. In classrooms where these beliefs are common, discussions are likely to be stifled as students will be less likely to present conjectures or partial arguments.

Throughout our study, the researchers strove to devolve responsibility for deciding whether a solution was correct to the students. A solution was only accepted if the class as a whole agreed that "it made sense". In Weber et al. (submitted for publication), we illustrate how this norm led students to carefully attend to, evaluate, question, and sometimes refine the arguments of others. Students' willingness to attend to and challenge one another's arguments was a necessary precondition for the observed discussion to create the learning opportunities described in this paper. In traditional classrooms, teachers are usually the arbiters of what solutions are correct. In such environments, we believe students will be unlikely to challenge their classmates' arguments, believing that it is the teacher's job to do so. We also suspect students will be less likely to carefully attend to their classmates' arguments, since it is not their responsibility to determine if the argument made sense.

Finally, in our study, students were given extended time to complete all of our tasks; they could work on an activity for as long as they saw fit. This may have encouraged the observed discussion in two ways. First, their extended experience with the Schoolopoly task and the inspection of the posters enabled them to form strong convictions and sensible rationales in the conclusions that they drew. Second, students were confident that they could debate for as long as they liked. In summary, we hypothesize that learning environments where student contributions are encouraged and not judged, sense making is encouraged and students are arbiters of what makes sense, and extended time is granted for investigations and discussion will invite students to attend to and challenge the arguments of others, which can make the warrants in students' discussion the objects of debate.

In our analysis, we did not focus on the teacher–researcher's role in this discussion. Instead, we focused on the cognitive benefits and learning opportunities of discussion and argumentation. The researcher's role was modest, primarily reinforcing the norms of the classroom that had been previously established. For instance, she asked students to share their conclusions and provide justifications for those conclusions and she reminded students to attend to one another's arguments. During the discussion, students sometimes became quite passionate and, at some points, many students were shouting their viewpoints simultaneously. When this occurred, the researcher would try to keep order by calling on a specific student to speak and asking that only one student speak at a time. In the beginning of the discussion, she would invite specific students to speak and occasionally summarize their contributions. At no point during the discussion did the researcher evaluate the students' arguments or push the direction of the discussion in a particular direction. For these reasons, we believe that the discussion analyzed in this paper was more a result of the environment that we created throughout our study than by specific actions by the teacher–researcher moderating the discussion.

Acknowledgments We would like to thank Chris Rasmussen, Matthew Inglis, and the reviewers for helpful comments on earlier drafts of this manuscript.

References

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

Bauersfeld, H. (1995). The structuring of structures: Development and function of mathematizing as a social practice. In L. P. Steffe, & J. Galle (Eds.) *Constructivism in education*. Hillsdale, NJ: Erlbaum.

Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education*, *28*, 258–277.

Drier, H. (2000). The *Probability Explorer:* A research based microworld to enhance children's intuitive understanding of chance and data. *Focus on Learning Problems in Mathematics*, *22*, 165–178.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Education Research Journal*, *30*, 393–425.

Inglis, M., Mejia-Ramos, J. P., & Simpson, A. (2007). Modelling mathematical argumentation: The importance of qualification. *Educational Studies in Mathematics*, *66*, 3–21.

Johnson-Laird, P. (2006). *How we reason*. Oxford: Oxford University Press.

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb, & H. Bauersfeld (Eds.) *The emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lampert, M., & Cobb, P. (2003). Communication and language. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.) *A research companion to principles and standards for school mathematics*. Reston, VA: NCTM.

Lampert, M., Rittenhouse, P., & Crumbaugh, C. (1998). Agreeing to disagree: Developing sociable mathematical discourse. In D. R. Olson, & N. Torrance (Eds.) *Handbook of education and human development*. Oxford: Blackwell.

Lubienski, S. T. (2000). Problem solving as a means towards "mathematics for all": An exploratory look through a class lens. *Journal for Research in Mathematics Education*, *31*, 454–482.

Manouchehri, A., & Enderson, M. (1999). Promoting mathematical discourse: Learning from classroom examples. *Mathematics Teaching in the Middle School*, *4*, 216–222.

McCrone, S. S. (2005). The development of mathematical discussion: An investigation in a fifth grade classroom. *Mathematical Thinking and Learning*, *7*(2), 111–133.

NCTM (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*, 23–41.

Radu, I., Tozzi, B., & Weber, K. (2006). Developing mathematical initiative in minority students. In J. Novotna, H. Moraova, M. Kratka, & N. Stehlikova (Eds.) *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, 4, 401–407. Prague, Czech Republic.

Rasmussen, C., & Stephan, M. (2008). A methodology for documenting collective activity. In A. E. Kelly, & R. Lesh (Eds.) *Design research in education*. Mahwah, NJ: Erlbaum (in press).

Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational Studies in Mathematics*, *46*, 13–57.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*, 114–145.

Stephan, M., & Rasmussen, C. (2002). Classroom mathematical practices in differential equations. *Journal of Mathematical Behavior*, *21*, 459–490.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference with probability simulation tools. *Journal of Mathematical Behavior*, *21*(3), 319–337.

Tarr, J. E., Lee, H. S, & Rider, R. (2006). When data and chance collide: Drawing inferences from empirical data. In G. Burrill (Ed.) *Thinking and reasoning with data and chance: 2006 yearbook of the NCTM* (pp. TBD). Reston: VA, NCTM.

Toulmin, S. (1969). *The uses of arguments*. Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning. *Psychological Review*, *90*, 293–315.

Weber, K., & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, *25*, 34–38.

Yackel, E. (2001). Explanation, justification, and argumentation in mathematics classrooms. In M. van den Heuvel-Panhuizen (Ed.) *Proceedings of the 25th Conference for the International Group of Psychology of Mathematics Education*, 1, 9–24, Utrecht, the Netherlands.