Computational Tools' Mediation of Argumentation in Undergraduate Probability and Statistics

Michael Foster, Rochester Inst. of Technology
Mary Lynn Reed, Rochester Inst. of Technology
Christian Cammarota, Rochester Inst. of Technology
Ben Zwickl, Rochester Inst. of Technology
Tony Wong, Rochester Inst. of Technology

As computation becomes increasingly central to mathematics education, instructors must balance competing forces when choosing which computational tools to use in their courses. This is compounded in probability and statistics where computation is widely used. Grounded in a social constructivist perspective, we believe that tools mediate our activities and that different tools play different mediational roles. As such, this study explores how different computational tools mediate undergraduate students' mathematical activity of argumentation. Using Toulmin's argument model, this research investigates how two classes in probability and statistics using different computational tools, R or Minitab, performed on a mirrored assignment. Through analysis of students' assignments, a difference emerged across the classes use of visuals. Our findings suggest Minitab promoted more deliberate consideration and use of visuals than R, leading to a difference in arguments produced by the students.

Keywords: Computation, Toulmin's Model of Argumentation, Statistics Education

The use of technology in undergraduate mathematics settings has routinely been emphasized over the last fifty years. This emphasis has been supported by technology expanding the set of possible activities for students and enhancement of their learning experiences (e.g., Chan et al., 2023). One such framing of technology is that of *computation*. Computation positions the use of specific kinds of technological tools as a means of mediating and augmenting problem solving.

DiSessa (2018) directly emphasized the impact computation can have on mathematics by drawing a parallel between computation and symbolic algebra. Since the invention of symbolic algebra, algebra has been integrated into society and has expanded the availability of complex mathematical understanding. Similarly, the promise and ubiquity of computational tools opens new opportunities for people to engage with and experience new mathematical understanding.

Indeed, many computational tools have been used in mathematics education to improve student understanding (Chan et al., 2023). In probability and statistics, in particular, there has been a consistent push for the use of computation to better augment and improve their learning experience, but also to prepare students for their future careers (Carver et al., 2016). Instructors are left to balance different considerations when choosing which tools to use in their courses. This work explores how different computational tools used in introductory courses influence students' mathematical activity.

Theoretical Framework

Our work is grounded in two theoretical frameworks: Vygotsky's social constructivism and diSessa's computational literacy. A central tenant of social constructivism is the claim that learning does not occur in isolation and, instead, happens in relation with others (Vygotsky, 1978). Beyond relations between other individuals, learning also occurs through relations between an individual and a culture, and a primary way this happens is with cultural artifacts or tools (e.g., a computer for the culture of American society). Significantly, all actions performed with a tool are transformed through the mediation of the tool. The higher quality the tool, the more influence it will have on our actions. According to Vygotsky (1987), this is characteristically illustrated by language, the quintessential tool of human culture.

For this study, we are interested in the broad set of cultural tools that leverage technology for problem solving, which we refer to as computational tools. Computational literacy (CL) is a framework that grounds our argument that computational tools are a cultural tool.

Computational literacy draws parallels between the use of computation and other representational forms that had, and continue to have, lasting effects on society (e.g., written language, algebraic symbols; diSessa, 2000). Framed in this way, computation shapes society and, as such, is a cultural tool. Computation can be explored through the material, cognitive, and social pillars of a literacy. Specifically, material CL encompasses the mechanisms of using computation (e.g., writing syntax); cognitive CL includes the application of computation to solve problems (e.g., using computation to visualize and solve differential equations); and social CL involves communicating through and about computation (e.g., the presentation of visualizations).

In summary, tools mediate our actions, different tools mediate differently, and computational tools are an important cultural tool. As such, the use of computation across society, and within mathematics specifically, raises several important questions about how computational tools mediate activities. Our research question guiding this project is as follows: *How do different computation tools mediate undergraduate students' mathematical activity of argumentation?*

Literature Review

Several projects have framed their inquiry of what we refer to as computation, broadly, under the exploration of technology in mathematics education (e.g., Biehler, 1997). Additionally, many projects have explicitly probed computation in mathematics education through the lens of computational thinking (e.g., Chan et al., 2023). Across both bodies of literature, there has been supporting empirical evidence for the use of technology to solve problems in mathematics.

While there is evidence supporting the use of computation at the undergraduate level—and in probability and statistics, in particular—there have also been contrasting claims about what those tools should look like. One school of thought centers on the opportunity to give students experience with technology that better prepares them for their future careers (National Science and Technology Council, 2023; Nolan & Temple Lang, 2010). For example, Nolan and Temple Lang (2010) proposed the importance of preparing introductory statistics students for the ever-evolving landscape of technology available to practicing statisticians. For these authors, this suggested that the tools being used should be versatile and have features that promote general computational skills (e.g., troubleshooting). On the other hand, many computational tools have been specifically designed for their use in introductory settings (e.g., TinkerPlots). For example, Biehler (1997) documented the design and possible uses of a tool called Modelling in connection with Exploratory Data Analysis and Stochastic Simulation (MEDASS). The goal of MEDASS was to provide a learning environment that leveraged the affordances of technology (e.g., for data analysis) in an easy-to-use environment. Ultimately, easy-to-use and versatile can be conflicting features. As such, instructors must balance the different affordances when choosing a tool for their introductory undergraduate courses (Johnson and Berenson, 2019).

To date, several projects have probed the affordances of different computational tools from either the expert or student perspective. In Abbasnasab Sardareh et al. (2021), four tools were compared (SPSS, basic R, Jamovi, and R Commander) across their usability (e.g., the presence of a graphical user interface [GUI]), technical features (e.g., variety of visualizations), and software capabilities (e.g., ability to manipulate data). Through the authors' assessment of these software, they concluded that tools with GUIs (e.g., SPSS) are important for introductory statistics, but acknowledged "[through] real software in [students'] introduction to statistics, they will be better positioned to use that software for real-world purposes later (p. 159)." In a similar study, Johnson and Berenson (2019) identified 11 criteria of computational tools that they had 13 faculty teaching introductory statistics use to rank a set of computational tools. JMP, a menu-driven statistical software, was the highest rated tool. These studies suggest that experts believe user-friendly tools like SPSS and JMP best serve their instructional goals for teaching statistics.

There have also been a handful of studies comparing students' uses of differing tools. Myint et al. (2020) compared students' uses and peer assessment of plots done in base R to those done using ggplot2, in R. Myint and colleagues reported that the users' of ggplot2 were more persistent and more likely to complete the assignments. Furthermore, through peer assessments, the plots produced in ggplot2 were perceived as clearer, containing more labels, and aesthetically pleasing. In another comparative study, Rode and Ringel (2019) measured students' anxiety and confidence about the use of R and SPSS outputs after they were exposed to either R or SPSS in their statistics course. Rode and Ringel reported no differences across the treatments in their pre/post-test changes in anxiety or confidence. Interestingly, both treatments evidenced reduced anxiety in using R *and* SPSS. These studies suggest that a difference can emerge in statistics courses that use different computational tools, but that is not always the case.

A question arises as to how differing tools influence specific mathematical activities (e.g., the activity of argumentation). As Wilensky (1995) demonstrated, computation can influence students' argumentation. As such, in this study, we narrow our exploration to the difference in students' argumentation mediated by different computational tools.

Methods

To compare how different computational tools mediated students' mathematical argumentation, a comparative design study was conducted. A homework assignment was designed to be completed by two classes in introductory probability and statistics at a university in the Northeastern United States. The assignment was co-designed by two of the instructors of record for the respective classes. As a separate part of their course design, the instructors had chosen computational tools that best served their instructional goals. In one class, students used Minitab (a menu-driven statistical software), and in the other, students used R (a programing language designed for statistical analysis) in Google Colab.

The primary goals of the assignments were for students to learn to use their respective tools, explore the gamma and Weibull distributions, and communicate their results. The focus of this report's analysis centers on the students' exploration of the parameters of the gamma distribution. In the gamma task, students generate random data sets and use histograms to visualize the gamma distribution. The goal of this task was for the students to manipulate the parameters, explore how they affect the distribution, and describe and justify their conclusions.

The assignments were designed to mirror each other, but differences in the tools required small variations in the assignments. Specifically, in the assignments the process of using the specific tool to generate data and visuals was described. In the Minitab assignment, the sequence of menus and windows was described. For Minitab, generating a random set of data fitting the gamma distribution begins with menu navigation (*Calc > Random Data…*). In the *Random Data* window, the user must enter a sample size in *Number of rows of data to generate* and a column name in *Store in column(s)*. Next, they must select the distribution they want the data to fit in the *Distribution* drop down menu (i.e., *Gamma*). Finally, they must enter values for the parameters in *Shape parameter* and *Scale parameter*. Once they select *OK*, a column of random data fitting the specifications of their gamma distribution will be generated under the chosen column name.

For R (in Google Colab), the analogous process can be done in a code cell using the function *rgamma()*. Thus, the R assignment began with a text block describing the *rgamma()* function and a linked resource describing the different parameters of the function (e.g., *n* for sample, *scale* for the scale parameter, *rate* for the rate parameter). Then, in a code block, scaffolded code was provided that required manipulation. This scaffolded code began with the assignment of a sample size, *n*. Students were told this through a code comment to "assign a number of samples to take." In the next line of code, a variable name is provided, "random_gamma_data," which takes the *rgamma()* function with input of *n*, and the two inputs of *shape* and *rate*. Students were prompted to set values for the latter two inputs, again, through code comments.

After following these steps in both assignments, students will have a set of random data fitting a gamma distribution either stored as a vector called "random_gamma_data" in R or as a column with a given name in Minitab. Next, both assignments described the process of turning that data into a visual in the form of a histogram. For Minitab, this again entailed instructions for the sequence of menus and windows to navigate. For R, scaffolded code, linked resource on the *hist()* function, and in code comments were used to guide students. After these instructions for using each tool, the assignments provided the same instructions, questions, and prompts.

Data and Analysis

Two data sources were used in this project. First, all students' work on the assignments were collected. In total, 68 artifacts were collected across the two classes (35 from Minitab users and 33 from R users) In addition, artifact-based interviews (Brennan & Resnick, 2012) were conducted with two students from each class. During the interviews, students were given their artifacts and asked to describe their work, thought processes, and documentation.

Analysis focused on students' work on the gamma distribution task because it was characteristic of students work throughout the assignment. Using Toulmin's model of argumentation (Toulmin, 2003), we coded for claims, grounds, and warrants. Claims, or conclusions, were centered on the effect the two parameters of the gamma distribution have on the distribution. Grounds are the ideas/data supporting the claim, and warrants provide the link between grounds and claims. The use of visuals as grounds emerged from our data. As such, artifacts were coded for the number of graphs included. This was further categorized into either no visual, one visual, or multiple visuals included. Open coding was used for interview data.

Findings

The following section is broken into two subsections. To begin, we present interview data showing that Minitab and R users' interpreted the goal of the task in the same way and engaged in a similar exploration process on the assignment. Next, we show the emergent differences within students' artifacts. This centers on an exploration of students' argumentation structure.

Students Goals and Processes

During the interviews, students were asked what they thought the goal of the gamma distribution task was. Both Minitab and R users claimed the goal was to use their respective tools to explore the distribution and its parameters. For example, one Minitab user said, "the main point of this task was … take those [distributions] and play with them, see what the different parameters do." Similarly, the other Minitab user explained, "[the goal was] predicting, trying to make a pattern … If I increase this parameter, do I expect the range to increase?" For the R users, one stated, "the main task was creating the gamma and Weibull distributions, changing the shape value and rate value, and seeing how those affect the distribution." The other R user said, “[the goal of the task was] to do a deep dive on this distribution, and you come up with your own ideas about what it means.” Consistently, students identified exploration as the goal of this task.

Next, during the interviews, students were probed about their exploration process and their choice to include visuals in their assignment. Both sets of students created multiple visuals that they did not include in their artifacts, but only Minitab users evidenced a consideration of which to include. For example, an R user who did not include multiple visuals in their artifact, described their process as, “just changing my histogram function over and over. Trial and error.” This user continued, “I didn't really, while I was going through, think to copy each iteration I had done or anything. So instead of leaving a trail, I was just rerunning the code every time I made the change.” This R user generated multiple visuals but explicitly did not think to leave “a trail” of them. Similarly, the other R user said “So it never came across my mind to show other graphs. I just thought, just change a value at a time, but then moved on.” While this R user included multiple visuals in their artifact, they claimed to not think about which of their visuals to include in their artifact. On other hand, a Minitab user, who included multiple graphs described a curation process in choosing visuals to include. They claimed, “I considered doing seven or five [visuals]. An odd number just to show a balance of decreased and increased parameters… Then I chose the three that made the most sense and, to me, display the most drastic changes.” Across the interviews, all students engaged in a similar exploration process of generating multiple visuals. Uniquely, Minitab users thought about which visuals to include in their artifacts.

Differences in Artifacts

A difference also emerged in the students’ artifacts. 34 (or 97%) of Minitab users included multiple visuals in their artifacts versus just 16 (or 48%) of the R users. For the remaining R users, 15 included one visual and 2 included no visuals.

The importance of the inclusion of multiple visuals was in the students’ ability to use the visuals as grounds for their claims. In the next two sections, we provide characteristic examples of the kinds of arguments contained in students’ artifacts. Namely, a difference emerged in arguments from students' artifacts that contained multiple visuals and those that did not.

Arguments using multiple visuals. When multiple visuals were included, the visuals served as grounds for students’ claims. Warrants supporting these grounds came in two forms. The first kind of warrant was a description of what the visuals illustrated. For example, one Minitab user included three graphs as *grounds* for each claim they made about the parameters. For the shape parameter, this user selected three graphs (Figure 1) and produced a warrant in their artifact:

> “In the three graphs above I varied the shape parameter to see how it affected the random data. It can be concluded that the shape parameter has an effect on the skewness of the data, as the [shape] parameter is increased the data becomes less positively skewed.”

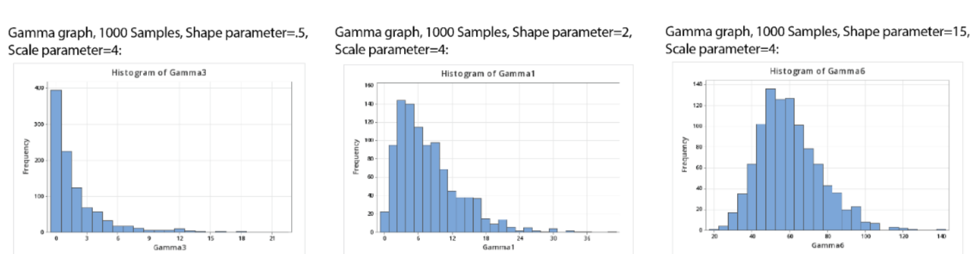

Figure 1. Three graphs generated by a Minitab user that served as grounds their claim.

This user made a *claim* that increasing the shape makes the skew “less positive.” This was supported by the three graphs (Figure 1), thus the graphs served as the *grounds*. Additionally, this user provided a *warrant* connecting the graphs to their claim by connecting the visuals to the effect they were illustrating (i.e., increasing the shape parameter caused a change in skew).

Another way in which the grounds of multiple visuals were supported by warrants was through examples. In one R users' artifact, they produced multiple visuals, including two with a rate parameter of 100 (Figure 2). They wrote, "At a very high rate the points on the graph are extremely close together. For example, when using the sample size of 100 with a rate of 100, all the values became less than 1." This users' *claim* was about the relationship between large values for the rate parameter and the spread of the data. The *grounds* for this user came in the form of four different visuals (including two visuals generated by the code in Figure 2). Then, a *warrant* of a specific example was used to illustrate how the visuals (i.e., the grounds) support the claim.

```
gamma_data3 <- rgamma(100, shape = 1, rate = 100)
hist(gamma_data3, breaks = 20)

gamma_data4 <- rgamma(100, shape = 50, rate = 100)
hist(gamma_data4, breaks = 20)
```

Figure 2. Segment of code from an R user's artifact

Arguments without multiple visuals. When users did not include multiple visuals, their arguments lacked grounds for their claims. Many of these users included the last visual they produced and, as the R user from the interviews described it, "moved on." This resulted in claims that made no reference to their visuals. As such, the visual did not serve as grounds. For example, one R user included one visual in their artifact and made two claims with no grounds:

> "As the value for shape goes towards infinity, the shape of the curve appears more like a bell curve… The peak of this curve also becomes centered on the x-value of the shape. So if shape = 1000, the curve will be centered on x = 1000."

This users' first *claim*, about skew, was accurate, but they did not provide grounds because the visual they included was of a gamma distribution with a shape parameter of 1 (i.e., right skewed). Their second *claim* was that the curve is "centered" on the shape value. This student then produced an example. Ultimately, this student did not include a visual of this example. As such, the example was grounds for their claim, but the lack of a visual of this example meant their argument did not contain a warrant.

In addition to providing visuals as warrants for claims, the inclusion of multiple visuals could have served to clarify the claims of many R users. One R user wrote, "Increasing shape decreases maximum value of the distribution." One interpretation of this response is that the student *claims* shape and range are inversely related. Alternatively, this user's *claim* was related to the normalizing effect of increasing the shape parameter. In either case, the use of visuals as grounds for this claim would have supported *and* clarified their claim. Similarly, several students made claims about the skew moving away from a right skew, but what the skew was moving towards was vague. One user wrote, "The gamma data starts skewed right when the shape value is low, but as I increase it the graph starts to become skewed further left." In the same vein, another user wrote, "When I change the shape parameter k, the skew-ness of the graph changes. When I have $k < 1$, it is skewed to the right, and when $k > 1$, it begins skewing more to the left." If these students' claims are to be taken as written, then these R users are making a *claim* about a left-skewed gamma distribution. This is not possible for a gamma distribution because its skew is strictly positive (i.e., right). Through the process of producing grounds for their claims (i.e., trying to find a visual that had a left skew) these users may have made an important discovery or honed their claims, as well as their arguments.

Discussion

In exploring how students' argumentation was mediated by the computational tool they used, it emerged that Minitab users put more consideration into their use of visuals and used more visuals. We then showed how these visuals were used in arguments by students as grounds with warrants for their claims. When students did not have multiple visuals, their arguments contained only claims. The production of justifications for claims was an explicit goal of this task. Thus, the differences in R and Minitab users' arguments, consideration, and number of visuals included suggests the tools were mediating their mathematical activity of argumentation.

While possible, we do not believe that the assignment or students' interpretation of the assignment played a mediating role. As described in the methodology, the assignments were designed to mirror each other as closely as possible. Furthermore, through our interviews, we identified that both R and Minitab users had viewed the goal of the assignment in the same way. This suggests that the assignment was not mediating the emergent differences.

We conjecture that the difference in each tools' ease of manipulation could explain the differences in the students' artifacts. In R, students can easily modify a segment of code and immediately observe changes (e.g., the effect of increasing the shape parameter). In contrast, Minitab requires users to navigate a sequence of menus to make a similar change. As an R user described, they were "changing [their] histogram function over and over" to quickly explore the distribution. This evokes a dynamic process of students looking at a visual, changing a parameter in their code, running their new code, and watching as the graph changes. Ultimately, this process leaves no "trail," or grounds that could be used to support claims. In Minitab, on the other hand, the process is not quick, and users must navigate a series of menus and windows. To see the dynamic effect changing the syntax produced for the R users, the Minitab users had to have multiple visuals produced and change tabs to view them. As such, the Minitab users had multiple visuals simultaneously accessible, and were able to include them in their final artifacts. In fact, this led them to pick which of their graphs "display[ed] the most drastic change."

Furthermore, when users produced their artifacts, R users were working in an environment that they exported as their artifact (i.e., they submitted an exported Colab Jupyter notebook). This means they could quickly go from seeing a pattern, writing about it in a text block, and moving on with the assignment. The Minitab users, on the other hand, could not produce text blocks in Minitab. As such, they needed to move their visuals to another document along with their written claims and justifications. It is possible that this extra step for Minitab users necessitated deeper consideration and curation of visualizations being presented.

When considering what computational tool to use in introductory courses, instructors must balance the desire to use a tool that students can easily leverage, one that meaningfully impacts their mathematical activities, and a whole host of other considerations specific to a given context (Johnson and Berenson, 2019). Importantly, different computational tools can mediate students' mathematical activities in different ways. Our findings suggest that the methodical process of exploration in Minitab produced better argumentation, and R's dynamic visualization emphasized exploration over argumentation. While both are productive activities, instructors must weigh their own goals when choosing what tool best serves their students. One suggestion we have is, consider a balanced approach between multiple computational tools. Leveraging a beginner friendly tool like Minitab can ease students into the use of computation and can illicit stronger argumentation structures. Then, integrating more powerful tools like R, can give students the opportunity to explore, dynamically, concepts that are easy to visualize in R.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Award No. DGE-2222337. Opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and are not reflective of the views of the National Science Foundation. We thank the members of the RIT Center for Advancing Scholarship to Transform Learning for fruitful conversations and continued support.

References

Abbasnasab Sardareh, S., Brown, G. T., & Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teaching Statistics*, *43*, S157-S172.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, *65*(2), 167-189.

Brennan, K., & Resnick, M. (2012, April). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association, Vancouver, Canada* (Vol. 1, p. 25).

Carver, R., College, S., & Everson, M. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*.

Chan, S.-W., Looi, C.-K., Ho, W. K., & Kim, M. S. (2023). Tools and Approaches for Integrating Computational Thinking and Mathematics: A Scoping Review of Current Empirical Studies. *Journal of Educational Computing Research*, *60*(8), 2036–2080. https://doi.org/10.1177/07356331221098793

diSessa, A. (2000). *Changing minds: Computers, learning, and literacy*. MIT Press.

diSessa, A. A. (2018). Computational Literacy and “The Big Picture” Concerning Computers in Mathematics Education. *Mathematical Thinking and Learning*, *20*(1), 3–31. https://doi.org/10.1080/10986065.2018.1403544

Johnson, M. E., & Berenson, M. L. (2019). Choosing Among Computational Software Tools to Enhance Learning in Introductory Business Statistics. *Decision Sciences Journal of Innovative Education*, *17*(3), 214–238. https://doi.org/10.1111/dsji.12186

Myint, L., Hadavand, A., Jager, L., & Leek, J. (2020). Comparison of beginning R students’ perceptions of peer-made plots created in two plotting systems: a randomized experiment. *Journal of Statistics Education*, *28*(1), 98-108.

National Science and Technology Council. (2023). *Building Computational Literacy Through STEM Education: A guide for federal agencies*.

Nolan, D., & Temple Lang, D. (2010). Computing in the Statistics Curricula. *The American Statistician*, *64*(2), 97–107. https://doi.org/10.1198/tast.2010.09132

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas* (Reprint). Harvester Press.

Rode, J. B., & Ringel, M. M. (2019). Statistical software output in the classroom: A comparison of R and SPSS. *Teaching of Psychology*, *46*(4), 319-327.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Vygotsky, L. (1987) *Thought and Language*. Cambridge, MA: The MIT press.

Wilensky, U. (1995). Paradox, programming, and learning probability: A case study in a connected mathematics framework. *The Journal of Mathematical Behavior*, *14*(2), 253–280. https://doi.org/10.1016/0732-3123(95)90010-1