IMPLEMENTING GAISE RECOMMENDATIONS THROUGH "DOING STATISTICS" TASKS

Brandon R. Hanson[1], Jeremy F. Strayer[2], Katherine A. Mangione[3], Jessica J. Brown[1], and Jeffrey D. Pair[1]
[1]Mathematics and Science Education Ph.D. Program, [2]Department of Mathematical Sciences, [3]Department of Elementary & Special Education
Middle Tennessee State University, Murfreesboro TN, USA
brh2g@mtmail.mtsu.edu

The Pre-K-12 GAISE report identifies four components of statistical problem solving: formulate questions, collect data, analyze data, and interpret results. Traditional tasks often require the teachers, not the students, to complete a significant portion of one or more of these components. To meet the need for students to engage in the complete statistical process, we have developed tasks that require students to make critical decisions for all four GAISE components. We term these doing statistics *tasks. This paper describes our implementation of* doing statistics *tasks with grades 7-12 in-service mathematics and science teachers during a professional development institute in the summer of 2013. We present an analysis of work samples that illustrate the ways in which* doing statistics *tasks enabled participants to engage in all four components of statistical problem solving.*

BACKGROUND

As human capacity to collect and analyze data has increased and as statistical analysis has become highly valued in multiple sectors of society, several national organizations have released curriculum recommendations that emphasize statistical literacy. The National Council of Teachers of Mathematics (NCTM) included data analysis and probability as one of their five content standards in the *Principles and Standards for School Mathematics* (NCTM, 2000). Four years later, the American Diploma Project's (2004) report on school mathematics curricula mirrored the NCTM objectives, and in 2007, the American Statistical Association endorsed the *Guidelines for Assessment and Instruction in Statistics Education (GAISE): A Pre-K-12 Curriculum Framework* (Franklin et al., 2007) in which the authors expanded upon the NCTM standards and clarified the components of statistical problem solving in the school curriculum. This trend of placing high value on statistical literacy continued with the publication of the Common Core State Standards in Mathematics (National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010) and the Next Generation Science Standards (NGSS Lead States, 2013). Of these documents, the GAISE report provides the most specific guidance for teaching statistics at the K-12 level. In particular, the GAISE report recommends that in addition to *studying* statistical processes, students must be given opportunities to *fully participate* in them. In particular, students must 1) formulate questions, 2) collect data, 3) analyze the data, and 4) interpret the results (Franklin et al., 2007).

In this paper, we describe how participants in the Statistical Reasoning and Thinking (StaRT) professional development project engaged in all four of the GAISE components while completing what we called *doing statistics tasks*. We describe the StaRT project and *doing statistics* tasks in detail and present the results of a qualitative study we conducted to answer the following research question: How do *doing statistics* tasks allow participants to substantively engage all four GAISE components?

THE START PROJECT

The yearlong StaRT professional development project held at Middle Tennessee State University was designed to help 7th - 12th grade mathematics and science teachers incorporate statistics-related content in their lessons while meeting the Common Core State Standards in Mathematics and the Next Generation Science Standards. Twenty-eight teachers participated in ten or more of these sessions. During the professional development the StaRT instructional team aimed to meet the following objectives:

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

1. Strengthen teachers' content knowledge in formulating statistical questions, collecting data, analyzing data, and interpreting data,
2. Develop teachers' pedagogical expertise in cultivating a standards-based learning environment as they teach statistics content using cognitively demanding and engaging tasks,
3. Enhance teachers' capacity to use technology effectively in math and science instruction, and
4. Support teachers as they implement, reflect on, and revise their own cognitively-demanding lessons.

In order to meet objectives 1 and 3, StaRT teacher participants completed *doing statistics* tasks and *conjecturing tasks* using the TI-Nspire graphing calculator and related technology tools. We have reported on the ways in which participants learned statistical concepts through *conjecturing tasks* (Strayer et al, 2013). To meet objectives 2 and 4, participants used the Common Core State Standards in Mathematics, Mathematical Task Framework (Stein, Smith, Henningsen, & Silver, 2009), Next Generation Science Standards, and GAISE (Franklin et al., 2007) to reflect on their experiences completing tasks in the StaRT sessions. Participants subsequently planned three of their own lessons to implement with their students in fall 2013. During the fall 2013 follow-up sessions, participants reflected on their lesson implementations and discussed some of the revisions that needed to be made in light of the successes and failures of the lessons to better align them with the above frameworks.

"DOING STATISTICS" TASKS

For K-12 teachers, conducting successful statistical investigations requires a complex set of skills. The GAISE document identifies four essential components of all statistical investigations ─ namely, formulating questions, collecting data, analyzing data, and interpreting results. The GAISE framework also includes three levels that define how each component should develop (A, B, and C). When formulating the question at Level A, for example, "teachers pose the question of interest" (p. 14) and the investigation is restricted to classroom application. However, when instruction progresses to Levels B and C, students "pose their own questions of interest" (p. 14) and questions may be generalized outside of the classroom application. These progressions are delineated for each of the four GAISE components (see pages 14 - 15 in Franklin et al., 2007). Traditional statistical tasks often do not allow students to engage all four GAISE components at high levels. To make tasks more "do-able" for students, teachers (or textbook authors) may complete one or more of the GAISE components (e.g., formulating questions and/or data collection) and ask students to focus on one or two components (e.g., data analysis and/or interpreting results).

We envision a different kind of task. *Doing statistics* tasks are unique in that they provide opportunities for students to substantially engage all four GAISE components at high levels. The structure is open and simple. To complete the task, the students will

1. Generate a list of questions they could ask about a given context and decide which of the questions on the list are statistical in nature,
2. Determine which of the questions could be answered in their classroom setting,
3. Decide what information they need in order to answer their question,
4. Decide how they will collect that information, and
5. Develop a plan for how they will use the collected information as a basis for answering their question.
6. Determine what the information they collected tells them about the question they are investigating and using pictures, symbols, and/or words to clearly communicate and support their conclusions.

The instructor guides students through the first five steps in small groups and/or as a whole class. Once students have a plan for collecting data, they get approval from the instructor to begin collecting the data. Students then complete the sixth and final step.

We named these tasks *doing statistics* to echo the characteristics of *doing mathematics* tasks conceptualized by Stein, Smith, Henningsen, and Silver (2009). In particular, *doing statistics* tasks require students to depend on their previous knowledge and experience. Students engage in all four GAISE components in a way that cannot necessarily be predicted, and certainly is not scripted. Furthermore, students must monitor their approaches and progress to determine if they are truly formulating an answer to the statistical questions they formulated. Lastly, because these tasks do not suggest a particular pathway toward solution, we expect that students would experience some level of anxiety (Stein, Smith, Henningsen, & Silver, 2009).

DOING STATISTICS: THE GUM TASK

In this section, we describe a *doing statistics* task we implemented with in-service teachers during one of the ten summer institute days of the StaRT professional development project.

General Description of the Task

The Gum Task was an adaptation of the By Golly, By Gum investigation from the Activities Integrating Math and Science (AIMS) Education Foundation's *Jaw Breakers and Heart Thumpers* materials (Wiebe, Ecklund, & Hilland, 1987, 1995). The original task provides students with a specific question about gum and progresses in a lock-step fashion that would best be described as a GAISE Level A investigation (Franklin et al., 2007, p. 14-15).

We adapted this AIMS lesson to encourage StaRT participants to work at higher GAISE Levels (Franklin et al., 2007, p. 14-15). Instead of the teacher posing the question about chewing gum mass and time chewed, we foster the six-step process by asking participants to answer the following questions:

1. What questions can we ask regarding bubble gum?
2. What questions can we answer today about bubble gum?
3. To answer our question(s), what information will we need?
4. How will we collect this information?
5. How might you use the collected information as a basis for answering your question?
6. What does your information tell you about the answer to your question? Use pictures, symbols, and/or words to clearly communicate and support your conclusions.

The Gum Task took nearly three hours to complete from start to finish. Teacher participants shared results by creating a poster that displayed their group's results and interpretations and presented the information on the poster to the class. We observed the implementation of this task and took field notes as participants completed the task. The posters and field notes were analyzed using open coding and a theme analysis (Corbin & Strauss, 2007) to determine how the *doing statistic*s task allowed StaRT participants to substantively engage all four GAISE components.

FINDINGS AND DISCUSSION

When completing the Gum Task, two StaRT participant groups investigated if there was a difference between how far two different gum brands can stretch. Two groups investigated how long it takes to chew different brands of gum until the gum is ready to blow bubbles of a pre-determined size (one or one and a half inches in diameter). Two groups investigated whether different brands produce larger bubbles. The final group investigated whether sugared gum loses mass faster than sugar free gum. In the sections that follow, we present the themes that emerged from our analysis of how this *doing statistics* task allowed in-service mathematics and science teachers to participate in the GAISE components.

Participants Engaged in All Four Components of Statistical Problem Solving

Of the seven participant teams, all but one demonstrated evidence of fully engaging in all components of the statistical problem solving process, albeit at different levels according to the GAISE framework. The negative case was a team that had formulated a statistical question, but got sidetracked while designing and carrying out the data collection. This team investigated how long it takes to chew different brands of gum until the gum could blow bubbles of 1.5 inches in diameter. Ironically, the team was trying to control for differences in bubble-blowing abilities between males and females by having one male and one female blow bubbles, but this data collection design adjustment unnecessarily introduced greater variability in the data. Being aware of potential sources of variability and crafting an experiment to design for these potential differences, is an indicator of engaging at higher levels in the GAISE framework, *but* the problem for this team was that their design changes were not anchored by their original question. The end result for this team was that they blocked by gender, collected some data, but were not sure how to connect it back to their original question. Not surprisingly, the team was unable to interpret the results of their analysis in a way that provided an informed answer to their original question.

The complications encountered by the above team suggest that when participants are designing their studies, they should ensure the data they collect will match the variables designated in their proposed statistical question. This *doing statistics* task asked participants to anticipate what kind of data they would need in order to answer their question, but this was asked long before they collected the data. It might be helpful to add an additional prompt that requires participants to reflect on their data collection design at the beginning of the data collection to consider whether the anticipated differences in the data they collect will allow them to draw conclusions about the question under investigation.

Participants Abandon Group to Group Comparisons When Distributions Overlap

Most participant teams appropriately chose boxplots to display the distributions of their data, but none of them specifically addressed the variability within each group (i.e. type of gum) compared to the variability between the groups. It appeared that two of the participant teams were making group to group comparisons because they noted that the entire distribution for one type of gum was higher than that of a second type. According to the GAISE framework, these participants were engaged at Level C for the “Analyze Data” component. However, three other groups that compared two types of gum collected data with distributions where some of the lower values of one group's distribution overlapped the higher values of the other group. Unlike the first two participant teams, these three teams struggled to interpret these overlapping displays. All three teams in this predicament failed to use distributional reasoning in favor of comparing centers or individual values to conclude that one group was “usually” more than the other. Because these teams were not using variability to make group to group comparisons, the GAISE framework would suggest that they were engaged at the lowest level (A) during the analysis phase. Yet, it seems unlikely that the classification of these five participant teams is this simple. Instead, we hypothesize that the two teams that made group to group comparisons did so because the differences were obvious. Had there been even slight overlap, they might have just as easily abandoned these comparisons in favor of other numerical measures that could be compared to come to an unambiguous conclusion. Similarly, the teams that had overlapping distributions might have made what would have appeared to be group to group comparisons had there been more of an obvious difference between the two groups.

Thus, the *apparent* level of engagement of a participant might have changed depending on the data they collected. We believe it is important to verify the actual level of engagement by asking hypothetical questions that would force those doing the task into more ambiguous settings to see if indeed they would maintain group to group comparisons or abandon them in favor of clear-cut numerical comparisons.

Numerical Measures are Not Necessarily Connected to Graphical Displays

Three of the six teams that used boxplots to analyze their data accompanied their boxplots with numerical measures. Every team that did this chose the mean as their measure of center and standard deviation as their measure of spread. This is problematic because the team's choice of numerical measure was disconnected from the displays of their distributions. This problem was compounded for teams that had distributions with extreme outliers that heavily skewed these non-resistant measures.

Because there was no apparent connection between the graphical displays and the reported numerical measures, we suspect that familiarity may have played a prominent role in the teams' data analysis decisions. Boxplots were likely chosen because they were more familiar to the participants than histograms or dotplots. Likewise, the mean is a more commonly used measure of center than the median, and standard deviation often accompanies the mean when technology is used to calculate statistical measures. Typically, reporting the interquartile range would require participants to subtract the first and third quartiles, making it a less attractive option if they are simply providing a numerical measure without considering its appropriateness.

This "complication due to familiarity" makes it difficult to pinpoint the data analysis GAISE level for these teams. Clearly, the participants were quantifying variability, and some were even making group to group comparisons. This suggests GAISE level C. However the almost parallel approach to analysis between teams that used conflicting graphical and numerical analyses suggests that even those that appear to be at Level C may not have engaged at such a high level.

Opportunity for Higher Levels of Engagement Depends on the Question

In the *doing statistics* Gum Task, most participants used pieces of gum as their observational units. During the data collection phase, they acknowledged the fact that there would be slight differences between pieces, but aside from observing as many units as time would allow, they did not express the need for randomization. This was, in part, because the questions they formulated did not inherently suggest the need for randomization, restricting their level of engagement to Level A. Some groups tried to incorporate controls of certain variables that would allow them to better focus on the differences between the two types of gum, but others did not. This occurred even though their design was constructed in a way that would allow them to collect data necessary to answer their question.

Like the data collection phase, the interpretation phase did not give participants an opportunity to engage in higher levels because there was little reason to suspect that the gum being observed was not representative of the gum population. Furthermore, because randomization was not utilized, they did not have the opportunity to reflect on its role in the results. Thus, there was no opportunity for these groups to engage beyond Level A. It might have been possible to encourage participants to address these higher GAISE level issues with specific prompts, but one must question whether or not students are truly engaged at a high level if they must be prompted to address these issues.

CONCLUSION

In this paper we presented *doing statistics* tasks as high cognitive demand tasks meeting all four GAISE components (Franklin et al., 2007). Specifically, the gum task as implemented in the StaRT summer institute provides an example of how such a task would meet the criteria for a doing mathematics task as described by Stein, Smith, Henningsen & Silver (2009) and would provide the opportunity for participants to engage in all four GAISE components. Also, we provided insight into how our group of practicing teachers approached these tasks and the levels at which they engaged in the each of the components. Finally, we highlighted some of the difficulties we encountered while trying to determine the level of engagement in each component using the GAISE framework.

By having the teachers complete the *doing statistics* tasks, we sought to raise their awareness and to develop their understanding of the different levels of the GAISE components so that they might create equally engaging tasks for their students. Furthermore, we provided them with a model lesson structure they could emulate in their own classrooms.

Although our work was with in-service teachers, we suspect these findings may carry over to middle and high school students. Future research should investigate this possibility and should study the role that the statistical knowledge for teaching plays in an instructor’s ability to facilitate a *doing statistics* task in a way that maximizes students’ opportunities to engage each component of the statistical problem-solving process at a high level.

To truly “do statistics”, one must engage in *all* components of the statistical process. Therefore, tasks such as the one described in this paper, must be designed so that learners have the opportunity to make high-cognitive demand decisions in each of these areas.

ACKNOWLEDGEMENT

The work presented in this paper was supported by a STEM Professional Development grant of *First to the Top* US federal funds administered by the Tennessee Higher Education Commission (2013).

REFERENCES

American Diploma Project (2004). *Ready or Not: Creating a High School Diploma that Counts.* Washington, D.C.: Achieve, Inc. Retrieved on June 19, 2010 from http://achieve.org/files/ADPreport_7.pdf

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd edition.). Thousand Oaks, CA: SAGE Publications, Inc.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics (GAISE) Education: A Pre-K-12 Curriculum Framework.* Alexandria, VA: American Statistical Association. Retrieved June 19, 2010 from http://www.amstat.org/education/gaise/

NGSS Lead States (2013). *Next Generation Science Standards: For States, By State*s. Washington, DC: The National Academies Press.

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

National Governors Association Center for Best Practices and the Council of Chief State School Officers (2013). *Common Core State Standards for Mathematics*. Washington, D.C.: National Governors Association Center for Best Practices and the Council of Chief State School Officers. http://www.corestandards.org/the-standards.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). Teachers College Press.

Strayer, J. F.., Hanson, B. R., Mangione, K., Pair, J., & Brown, J. (2013). Using conjecturing tasks to support in-service teachers' conceptual statistical knowledge. In *JSM Proceedings*, Section on Statistical Education. Alexandria, VA: American Statistical Association.

Wiebe, A., Ecklund, L., & Hillen, J. (Eds) (1987, 1995). *Jaw breakers and heart thumper*s. Fresno Pacific University: AIMS Educational Foundation.