# STRATEGIES EMPLOYED BY SECONDARY MATHEMATICS TEACHERS ON INFERENTIAL REASONING TASKS

Maryann E. Huey[1] and Wendy Weber[2]
[1]Drake University, Des Moines, Iowa, United States of America
[2]Central College, Pella, Iowa, United States of America
maryann.huey@drake.edu

*This paper describes strategies employed by 49 secondary mathematics teachers on inferential reasoning tasks with special attention to randomization approaches. The participating teachers engaged in a year-long professional development program and experienced activities spanning the secondary content standards specific to probability and statistics widely adopted throughout the United States including randomization approaches to inferential tasks. Pre- and post-assessment data were collected to gauge changes in teachers' content knowledge from two sources: GOALS-2 and five research-based, constructed-response items. We characterize the strategies teachers demonstrated on post-assessment inferential reasoning tasks and note changes in strategies from pre- to post-assessments. Lastly, we discuss implications for professional development efforts specific to secondary mathematics teachers.*

## INTRODUCTION

With the recent emphasis of statistics as a core component of the middle and secondary mathematics curriculum in the United States and other countries, the role of inference is gaining in prominence (Senior Secondary Board of South Australia [SSABSA], 2002; Ministry of Education and Sciences, 2007; NGA & CCSSO, 2010). Current recommendations for middle and secondary statistics educators outlined in the *Guidelines for Assessment and Instruction in Statistics* (*GAISE*) *Report*: *A Pre-K-12 Curriculum Framework* advocate introducing students to informal inferential reasoning tasks during middle school with formalization occurring during secondary years (Franklin et al., 2007). These recommendations are evident in the articulation and sequencing of the Common Core State Standards for Mathematics (CCSS-M) adopted throughout the United States (NGA & CCSSO, 2010) with simulation and randomization approaches emphasized during secondary schooling. While programmatic recommendations are evolving, such as *The Statistical Education of Teachers* guidelines (Franklin et al., 2015), little is known about how teachers learn to inferentially reason with simulation and randomization approaches. In this paper, we report upon the changes in teachers' strategies specific to inferential reasoning tasks after an intensive professional development experience.

## METHODOLOGY

Two cohorts of teachers are included in this study for a total of 49 participating secondary mathematics teachers. The cohorts represent a variety of prior educational and teaching experiences related to statistics and probability. In terms of prior coursework, 7 teachers reported no prior coursework in statistics or probability, 30 reported one prior content course, 10 reported two prior courses, and 2 reported taking three or more classes. Of the prior courses taken, 43 were undergraduate content courses, 8 were master's level content courses, and 2 were specifically aimed at the pedagogy of teaching statistics and probability.

In terms of regular teaching responsibilities, 15 participating teachers reported never teaching statistics or probability, 14 reported teaching basic probability concepts as part of another content course, 6 reported teaching some statistics as part of an algebra course, and 13 reported teaching statistics regularly. Of those with regular teaching responsibilities, 4 taught advanced placement statistics, a college-level course offered in the United States and Canada and overseen by the College Board, and 9 taught a college-equivalent introduction to statistics course.

Each cohort completed a year-long professional development program including both content and pedagogical recommendations for teaching secondary statistics and probability in alignment with the CCSS-M. We allocated approximately one-third (20 hours) of our content-specific instructional time to simulation and randomization approaches with inferential reasoning tasks, employing technology as well as manipulatives (e.g., rolling dice, shuffling cards). Formal hypothesis testing was not included in the professional development program. Primary sources utilized during this instruction included: 1) STEW, an online repository of peer-reviewed lesson plans aligned with the CCSS-M sponsored by the American Statistical Association and National Council of Teachers of Mathematics (NCTM), 2) "Navigating through Data Analysis in Grades 9-12" (NCTM, 2003), and 3) "Introduction to Statistical Investigations" (Tintle, et al., 2015).

Two pre- and post-assessments were administered to all participating teachers: the Goals and Outcomes Associated with Learning Statistics 2 (GOALS-2) (Sabbag & Zieffler, 2015) assessment and a constructed-response assessment consisting of five research-based, open-response items. While both assessments captured gains in teachers' content knowledge in relation to the professional development, the changes in teachers' strategies were evident on the constructed-response assessments. The five-item assessments consisted of one task focused on interpreting the results of a simulation, one sampling task, one probability task, and two inferential reasoning tasks. We focus here on the two inferential reasoning tasks, unchanged from pre- to post- assessment, as these items were ill-structured in nature and required participating teachers to choose an approach rather than follow a prescribed method.

The two items will be referred to by their names, *Speed Trap* and *Review Session.* The *Speed Trap* item was adapted from Cobb (1999) and requires comparing two quantitative samples of size 60 obtained before and after a vehicle speed trap was implemented. Teachers were asked to say whether a speed trap was successful or not based upon visual inspection of data distributions and any analyses that they choose to perform. The *Review Session* task is an adaptation of a task created by Zieffler et al. (2007). We provided population and sample data to participating teachers and asked them to determine whether a review session improved students' performances on an exam. This task required participating teachers to compare a sample's characteristics to population data. For both tasks, we provided dot plot representations of the data, the raw data files, posed open-ended questions, and required justification. Through this structure, participating teachers could choose to reason informally, perform formal hypothesis tests, or engage in randomization approaches.

The authors separately coded the responses to these two items as either informal, formal, or randomization. According to Garfield and Ben-Zvi (2004), informal inferential reasoning responses are characterized by coordinating relevant statistical ideas and concepts, such as central tendency, spread, and the shape of the data distributions, in order to generate a conclusion that extends beyond the data at hand. Formal approaches, for purposes of this paper, are characterized by application of a hypothesis test coupled with an accurate interpretation of results and an inferential statement. Lastly, randomization approaches are characterized by obtaining a collection of simulated statistics, assuming the null hypothesis is true, and then again interpreting the results in order to generate an inference. Once all pre- and post-assessment responses to these two items were coded, the authors compared ratings and obtained an inter-rater reliability of 99%.

In addition to choice of strategy, we coded the responses within each category type as either unproductive, productive with one statistical idea or aspect, productive with multiple statistical ideas or aspects, or relational, meaning a fully integrated and comprehensive response incorporating all key ideas. For the cases of one and multiple productive statistical ideas, responses may have included uncoordinated or unproductive ideas in addition to the productive element or elements identified. A response identifying a general approach including what would be compared, but without the comparison completed, was coded as productive with one statistical idea. The authors created a rubric with detailed criteria for the designated levels and then coded pre- and post-responses for both items separately across all participating teachers. Upon comparison, an inter-rater reliability of 76% was achieved with all discrepancies resolved through discussion.

## RESULTS

*Speed Trap.* Overall, participating teachers consistently pursued informal inferential reasoning strategies for the *Speed Trap* item on the pre- to post-assessments. Only one participating teacher employed a formal hypothesis test on both assessment responses, and ten teachers attempted a randomization strategy on the post-assessment for this item. See Table 1 below.

Table 1. *Speed Trap* pre- and post-assessments changes

| Pre- / Post- | Informal | Formal |
|---|---|---|
| Informal | 38 (78%) | 0 (0%) |
| Formal | 0 (0%) | 1 (2%) |
| Randomization | 10 (20%) | 0 (0%) |

On the post-assessment, approximately 53% of the responses were coded as either multiple statistical ideas or relational, indicating productive inferential reasoning skills across all strategy choices. For the subset of randomization responses, 60% of responses were coded as either multiple or relational. We share one of the two relational responses to highlight a fully coordinated and thorough response for a randomization approach. “I performed a randomization test. I wanted to see if it was statistically significant that the mean difference was 1.575 mph. Using Rossman and Chance two-means randomization, I shuffled and counted how many times by chance the mean difference was 1.575 or greater. This occurred 515/50,000=1.03%. Since my p-value is small, I can say it was unlikely to see my sample mean by chance. There is strong evidence that the speed trap effectively reduced speed”.

Of the four participating teachers with multiple productive statistical ideas, responses tended to include the correct choice of applet, a reasonable result, but challenges either interpreting the results or generating an inferential statement. Of those that were coded as productive with one idea, participating teachers tended to report that they wanted to complete a simulation but did not know how to either choose the right applet or enter the data into the applet. Others reported results that demonstrated an incorrect approach was selected. For example, one teacher ran a matched-pairs simulation for this task rather than assuming the two samples were independent. Another participating teacher combined all the sample data and then generated random samples from the combined data. Two others noted reasonable applet choices but could not enter the data correctly into the applet.

*Review Session.* Similar to the *Speed Trap*, participating teachers again embarked upon informal inferential reasoning strategies from pre- to post-assessment but to a lesser extent for this item. Four teachers employed a formal hypothesis test on the pre-assessment with one switching to randomization on the post-assessment. Seventeen teachers switched from informal approaches on the pre-assessment to a randomization strategy on the post-assessment. See Table 2 below.

Table 2. *Review Session* pre- and post-assessments changes

| Pre- / Post- | Informal | Formal |
|---|---|---|
| Informal | 27 (55%) | 0 (0%) |
| Formal | 1 (2%) | 3 (6%) |
| Randomization | 17 (35%) | 1 (2%) |

On the post-assessment, approximately 53% of item responses for the *Review Session* were coded as either multiple statistical ideas or relational. For the subset of randomization responses, 53% of responses were also coded as either multiple or relational. We share one of the five relational responses to highlight a fully coordinated and thorough response for a randomization approach. “Sampling 75 students from the population 100,000 times resulted in 2807 (2.81%) of the time getting this group’s average score or better. So, we can conclude the review made an impact”. For responses coded as multiple or relational, teachers generally completed a randomization fully with correct interpretations or completed randomizations with a small amount of trials, such as generating 100 samples rather than 1,000 or greater. Of the responses coded as productive with one idea, participating teachers tended to recommend that a randomization be completed to determine whether the results were significantly different for the review session sample.

## DISCUSSION AND IMPLICATIONS

Researchers have found that mathematics teachers in the United States feel least prepared to teach statistics and probability content (CBMS, 2012) and that changes in actual teaching practices often lag content recommendations (Jones & Tarr, 2010). Generally speaking, teachers often omit or provide superficial treatment to content that they are uncomfortable with themselves. Encouragingly, many participating teachers shifted responses from informal to randomization approaches on the pre- and post-assessment responses driven by a need to quantify *significant differences*. In addition, the success of participating teachers in utilizing randomization approaches was the same or better than informal and formal approaches. However, the dominant trend in responses remained with informal methods from pre- and post-assessment. This finding seems to imply that teachers were most confident in their abilities to informally reason about core statistical ideas, which is a desirable attribute for secondary teachers of mathematics. However, statistical significance cannot be accurately quantified through such approaches. Further, teachers require knowledge of randomization approaches in order to effectively teach statistical inference as outlined by the CCSS-M (NGA & CCSSO, 2010). From our study, it appears that a portion of participating teachers gained the needed confidence and skills to engage in randomization approaches on inferential tasks, but others require additional experiences.

## REFERENCES


Burrill, G., Franklin, C., Godbold, L. & Young, L. (2003). *Navigating through data analysis in grades 9-12*. Reston, VA: National Council of Teachers of Mathematics.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*(1), 5-43.

Conference Board of the Mathematical Sciences. (2001). *Issues in mathematics education volume 11: The mathematical education of teachers.* Providence: American Mathematical Society.

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R. & Spangler, D. (2015). *The statistical education of teachers.* Arlington, VA: American Statistical Association.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework.* Alexandria, VA: American Statistical Association.

Garfield, J. & Ben-Zvi, D. (2004). Statistical literacy, reasoning and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-15). Dordredcht, The Netherlands: Kluwer.

Ministério da Educação (2006). *National curricular guidelines for secondary school: Nature sciences, mathematics, and their technologies.* Brasilia, Brazil: Author.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common core state standards: Mathematics standards.* Retrieved July 17, 2017 from http://www.corestandards.org/Math/

Sabbag, A. G. & Zieffler, A. (2015). Assessing learning outcomes: Analysis of the GOALS-2 instrument. *Statistics Education Research Journal, 14(2), 93-116.*

Senior Secondary Board of South Australia. (2002). *Mathematical studies curriculum statement.* Adelaide, Australia: Author.

Statistics Education Web (STEW). Retrieved November 14, 2017 from http://www.amstat.org/ASA/Education/STEW/home.aspx

Tarr, J E., & Jones, D. L. (2010). Recommendations for Statistics and Probability in School Mathematics Over the Past Century. In B. J. Reys, R. E. Reys, & R. Rubenstein (Eds.), *Mathematics curriculum: Issues, trends, and future direction: Seventy-second yearbook* (pp. 65-76). Reston, VA: National Council of Teachers of Mathematics.

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T. & VanderStoep, J. (2015). *Introduction to Statistical Investigations*. Hoboken, NJ: John Wiley & Sons, Inc.

Zieffler, A., delMas, R., Garfield, J., & Gould, R. (2007). Studying the development of college students' reasoning about statistical inference. Paper presented at the *Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.*