COGNITIVE TRANSFER ASSESSMENT IN POST-SECONDARY STATISTICS

Matthew D. Beckman and Robert delMas
Pennsylvania State University, USA
University of Minnesota, USA
mdb268@psu.edu

This mixed-methods study describes learning outcomes related to cognitive transfer among post-secondary students upon completion of a first course in statistics and makes a case for emphasis of such outcomes as part of the curriculum. A new assessment tool called the Introductory Statistics Understanding and Discernment Outcomes (ISTUDIO) assessment tool was developed and administered to nearly 2000 students attending a wide variety of post-secondary institutions primarily in the United States. Students were observed to demonstrate measurable evidence of both forward-reaching and backward-reaching transfer outcomes. Unexpected response patterns revealed by the analysis of constructed response tasks are presented. Implications for teaching and research are discussed with attention toward both simulation-based and non-simulation-based introductory statistics curricula.

INTRODUCTION

Statistical thinking in part concerns comprehension of "how, when, and why" a statistical framework can inform some inquiry (Ben-Zvi & Garfield, 2005). In learning and cognition research, an important mechanism by which students accomplish this sort of comprehension is sometimes referred to as transfer. Transfer can be defined as the degree to which acquired knowledge is applied to new or novel situations (Singley & Anderson, 1989).

The purpose of knowledge transfer can be distinguished based on whether knowledge is generalized for an undetermined future use—forward-reaching transfer—or whether available knowledge/experience is applied to a task at hand—backward-reaching transfer (Salomon & Perkins, 1989). Wild and Pfannkuch (1999) describe a similar notion in which archetypical mental models are abstracted that may inform problem solving in the context domain. Wild and Pfannkuch (1999, p. 244) explain "there is a continual shuttling between the two domains and it is in this shuttling or interplay, that statistical thinking takes place."

Regardless of the intended type of transfer, successful transfer requires intentional effort (Bransford, Brown, & Cocking, 2000; Singley & Anderson, 1989). Students without explicit intervention will struggle or fail to transfer even when problem sets are extremely similar (Cooper & Sweller, 1987; Singley & Anderson, 1989). Similar results have been observed in the context of statistics education (Garfield, delMas, & Zieffler, 2012; Lovett & Greenhouse, 2000).

While there are approaches to assess propensity to transfer knowledge, no published assessment existed to measure transfer outcomes for students of introductory statistics. The goal of this paper is to introduce and summarize results of a larger study (i.e., Beckman, 2015) in which the Introductory Statistics Understanding and Discernment Outcomes (ISTUDIO) instrument was designed to quantify transfer outcomes for diverse introductory statistics curricula.

METHODS

A test blueprint of the structural organization of the ISTUDIO instrument was drafted and revised. Candidate tasks were then developed and critiqued based on criteria dictated in the test blueprint. Expert feedback and cognitive interviews with student participants were used to improve instructions, tasks, and structural considerations prior to finalizing the instrument for large-scale field testing. Instructors were recruited to participate in the field test through various methods in order to seek diversity in curriculum design and instructional practice. Students completed the ISTUDIO instrument online at or near the end of an introductory statistics course. Data analysis included evaluation of the reliability and validity of the instrument as well as an abbreviated item analysis. Additionally, qualitative analysis of expert feedback provided insight about the contribution of the instrument for curriculum development decisions.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

I-STUDIO Development

The measured construct was defined as the ability to transfer conceptual understanding of statistical inference to novel problem settings. This construct was thought to require both the ability to identify novel problem scenarios that warrant application of statistical inference, and the ability to achieve forward-reaching and backward-reaching transfer of statistics knowledge. Backward-reaching transfer tasks required discernment of whether a context would or would not benefit from application of statistical inference and demonstration of an appropriate solution strategy. Students were encouraged to frame solutions as though they were giving advice to a classmate rather than producing computations or formulas. Forward-reaching transfer tasks described a conceptual model and then asked students to generate an appropriate context and map specific components of the conceptual model to the chosen context.

The assessment was expected to include primarily open-ended tasks. To reduce cognitive load, the initial framework called for each primary outcome to be assessed by two or three contexts (e.g. data sets, data stories, etc.) with one or more open-ended prompts accompanying each context. Each open-ended prompt was considered to represent an item. Before items were written, a test blueprint was developed and refined with iterative feedback from expert reviewers in order to define each specific item type and dictate the number of items allocated to each type.

The final ISTUDIO assessment may be requested by email from the first author, however descriptions of a couple example items are included here as an illustration. One ISTUDIO forward-reaching transfer task is based on an item described by Chance (2002) and credited to Rossman and Chance (2001):

> The underlying principle of all statistical inference is that one uses sample statistics to learn something… about the population parameters. [Write] a short paragraph describing a situation in which you might use a sample statistic to infer something about a population parameter. Clearly identify the sample, population, statistic, and parameter in your example. [Be specific,] and do not use any example which we have discussed in class.

Another ISTUDIO backward-reaching transfer task is similar to one described by Garfield et al. (2012, p.898):

> Some people who have a "good ear for music" can identify the notes they hear when music is played. One note identification test consists of a music teacher choosing one of the seven notes (A, B, C, D, E, F or G) at random and playing it on the piano. The student is standing in the room facing away from the piano so that he cannot see which note the teacher plays on the piano. The note identification test has the music student identify 10 such notes.

In the ISTUDIO assessment, a task such as the latter was presented to the student with two accompanying prompts: (a) "Should statistical inference be used to determine whether the student has a "good ear for music"? Explain why you should or should not use statistical inference in this scenario" and (b) "explain how you would decide whether the student has a good ear for music using the note identification test. *(Be sure to give enough detail that a classmate could easily understand your approach, and how he or she would interpret the result in the context of the problem.)*" Part (a) evaluates the ability of students to discern whether or not a described scenario would benefit from statistical inference, and then outline a productive solution strategy in part (b). For completeness, ISTUDIO also included two items for which inference was not appropriate because of complete access to the population or knowledge of a deterministic process.

Field-Testing I-STUDIO

The purpose of the large-scale field test was to obtain a diverse sample that allowed the I-STUDIO instrument and rubric development to be tested across a variety of students, courses, and use scenarios. In order to solicit participation of introductory statistics instructors, announcements were broadcast through the Isolated Statisticians of the American Statistical Association (ISOSTAT), American Statistical Association (ASA) Section on Statistical Education, and Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), list serve outlets. These list serves were expected to provide access points to a nation-wide target audience.

The ISTUDIO assessment was administered using a web-based interface by fourteen (14) instructor participants representing a total of 29 class sections for 16 unique courses at 15 institutions. Submissions were anonymous and the interface was configured to prohibit multiple responses from the same IP address (i.e., student) A total of 1975 students submitted responses, among whom 1935 consented to participate in the research study. In order to restrict the data to earnest attempts, attempts submitted in fewer than 10 minutes and incomplete responses were omitted from the final data set. The resulting sample size included 1566 unique student participants. Maximum course enrollment aggregated across all participating institutions was estimated as 2265, indicating a total response rate of 87% and total usable response rate of 69%. These percentages may underestimate actual response rates as they assume that every course ended the semester with maximum enrollment.

Since there were more responses than anticipated, a representative sample was selected for scoring and analysis. The sampling method randomly chose 12 students from each course; two small courses had fewer than 12 responses submitted, so all were included in those cases. The resulting sample included 178 respondents representing 16 different post-secondary introductory statistics courses.

RESULTS

Summary of Psychometric Analysis

While detailed psychometric analysis is beyond the scope of this paper, a detailed analysis of scoring consistency, reliability, and validity was reported with the larger study in Beckman (2015). Scoring consistency data showed greater than 90% inter-rater and intra-rater agreement. Spearman-Brown split-half reliability was used since the assessment tool measures a multi-dimensional construct; a 95% confidence interval for Spearman-Brown reliability was [0.71, 0.85]. The study adopted a unified view of validity and pointed to favorable evidence from multiple expert and peer reviews of the test blueprint and assessment tool, scoring consistency, reliability, and confirmatory factor analysis. Confirmatory factor analysis provided statistical evidence for the distinction between forward-reaching and backward-reaching transfer, though “discernment” ability was not statistically significant.

Noteworthy Response Patterns

Qualitative analysis of exceptional responses was focused primarily on noting emergent understanding. Items on each dimension offer unique insights that may not be easily observed with a unidimensional or forced-response assessment tool. For example, a frequent pattern among the responses revealed a misconception that a population for the purposes of inferential statistics should or must be a population of physical people or objects rather than a population of outcomes for some process. Consequently, corresponding solutions for items with a context describing a process commonly imposed an artificial population of people for the purpose of inferential comparisons rather than a population of outcomes, as expected.

Another noteworthy theme was the prevalence of contradictory responses among backward transfer “discernment” tasks that require students to (a) determine whether or not statistical inference is necessary before recommending (b) a problem-solving method. Students frequently advocated for statistical inference in (a) and then described a non-inferential solution in (b) or rejected the need for inference in (a) and described an inferential strategy in (b).

Finally, among the forward-reaching transfer items, the data revealed remarkable variability explicitly stated to represent the parameter of interest in a scenario of their own invention. Responses were observed to conflate the parameter with almost every other detail of the scenario including details of study design, populations, variables, and more.

CONCLUSIONS

Evidence indicates that the I-STUDIO instrument measures both forward-reaching and backward-reaching transfer outcomes related to statistical thinking with strong psychometric properties. Prior to this study, no published assessment had been designed to measure this specific outcome. The results may support a theory posed by Wild and Pfannkuch (1999) in which statistical thinking takes place in the act of mental shuttling back and forth between the context domain and the schema for abstract modeling archetypes. The evidence of both forward-reaching and backward-reaching transfer as distinct dimensions suggests an ability to isolate and measure the dexterity with which students perform as they shuttle in each direction.

With strong reliability and validity evidence supporting its use as an instrument to measure cognitive transfer outcomes associated with the introductory statistics curriculum, the ISTUDIO instrument may fill an important assessment role for the statistics education community to inform reliable comparisons of transfer outcomes.

Limitations

The sample of participating courses apparently did not represent quite as much curriculum diversity as anticipated. All students that participated in the cognitive interviews had completed a course with at least half of its curriculum devoted to simulation-based methods, though this approach was not well-represented during the field test. While diverse responses were well-accommodated by the scoring rubric and provided preliminary evidence toward the aim of designing the assessment tool with robustness to curriculum approach, more work is needed. Accompanying the field test data, no demographic data were available to assess differential item functioning, and final grades (or expected grades) were not available to corroborate ISTUDIO scores as validity evidence.

Implications for Teaching and Research

If transfer outcomes are of value for the introductory statistics curriculum, then the ISTUDIO assessment tool provides an instrument with good psychometric properties that teachers can use for comparing outcomes of alternative curricula. Additionally, the I-STUDIO instrument can be used to measure the effect of curriculum changes designed to improve transfer outcomes. Again, the instrument and rubric are designed with intent to accommodate diverse curricula for the purpose of evaluating course outcomes. Furthermore, the variability explicitly attributed to the parameter of interest in a context of students' own choosing prompted the authors to conduct a follow-up study of emergent understanding related to statistical modelling using an independently drawn random sample of 500 student responses, which has been submitted for publication elsewhere.

REFERENCES

Beckman, M. D. (2015). *Assessment of cognitive transfer outcomes for students of introductory statistics* (Doctoral dissertation, University of Minnesota—Twin Cities). Retrieved from http://iase-web.org/documents/dissertations/15.MatthewBeckman.Dissertation.pdf

Ben-Zvi, D., & Garfield, J. (2005). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-15). Springer, Dordrecht.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: National Academies.

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3).

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education, 44*(7), 883-898.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician, 54*(3), 1-11.

Rossman, A. J., & Chance, B. L. (2001). *Workshop statistics: Discovery with data (2nd ed.).* Emeryville, CA: Key College Publishing.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist, 24*(2), 113-142.

Singley, M. K., & Anderson, J. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-248. doi:10.1111/j.1751-5823.1999.tb00442.x