Leveraging Cognitive Theory to Create Large-Scale Learning Tools

Darryl Chamberlain Jr.
University of Florida

Russell Jeter
Emory University

At the 21st Annual Conference on Research in Undergraduate Mathematics Education, Ed Dubinsky highlighted the disparity between what the research community knows and what is actually used by practicing instructors. One of the heaviest burdens on instructors is the continual assessment of student understanding as it develops. This theoretical paper proposes to address this practical issue by describing how to dynamically construct multiple-choice items that assess student knowledge as it progresses throughout a course. By utilizing Automated Item Generation in conjunction with already-published results or any theoretical foundation that describes how students may develop understanding of a concept, the research community can develop and disseminate theoretically grounded and easy-to-use assessments that can track student understanding over the course of a semester.

Keywords: Assessment, Automated Item Generation, Technology in Mathematics Education

At the 21st Annual Conference on Research in Undergraduate Mathematics Education, Ed Dubinsky highlighted the disparity between what the research community knows and what knowledge is actually applied by practicing instructors. This disparity is not unique to mathematics education and exists even in education research in general (Van Velzen, 2013). One method to bridge this disparity is to involve mathematics instructors in research projects (Vidakovic, Chen, & Miller, 2016). Even outside of funded efforts, the RUME community in general has been encouraging practicing instructors to participate. However, there are a plethora of instructors that this approach cannot be applied to as they have limited time and resources. For these instructors, administering and providing feedback for open-response classroom activities is not feasible, especially for those who coordinate large-scale courses such as Calculus. This burden renders the knowledge of the research community moot, as it is too resource-expensive to collect and implement the knowledge practically. We propose an alternative method to engage these instructors: utilize the mathematics education literature and programming languages to dynamically generate multiple-choice questions that can form easy-to-use assessments throughout a course. To set the stage for this method, we briefly summarize how research has been presented to the community at large.

Disseminating Research Results

Let us consider how mathematics education research is disseminated from the perspective of a mathematics instructor. First, we need access to journal articles on the results from research - articles that may or may not be available through our school's library. Assuming we get access, we read through the results and notice the majority of assessments used are free-response assessments. This is not unexpected as qualitative research primarily use free-response assessments to gain as much insight into student thinking as possible. However, these results and assessments are difficult for instructors to utilize. The result is an isolating effect where instructors rely on their own experience and knowledge to develop instruments that may or may not be based in how students develop their mathematical conceptions.

One avenue for potential instructor use of research results is *concept inventories* - multiple-choice assessments designed to explore students' conceptual knowledge of a specific topic. One of the earliest such assessments is the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992), which outlines the conceptions necessary to understand Newtonian force. More recently in mathematics education, Carlson, Oehrtman, and Engelke (2010) introduced the Precalculus Concept Assessment (PCA). We quickly summarize how they developed this assessment below.

Through numerous studies, Carlson et al. (2010) developed a taxonomy of the necessary conceptions students should develop before taking calculus. Each assessment item is linked to one or more of these conceptions and went through multiple phases of refinement and validation:

- Phases I & II: Series of studies to identify and analyze how students understand the central ideas of precalculus and calculus. Open-ended questions were refined and common student responses were identified.
- Phase III: Validated multiple-choice items based on the open-ended questions from Phases I & II. This phase went through eight cycles of administering the assessment, conducting follow-up interviews, analyzing student work, and revising the taxonomy and assessment based on the results.
- Phase IV: Widespread administration of the revised 25-item multiple-choice assessment.

Based on this short explanation of their generation and validation process, it is no wonder few concept inventories have been developed to date – these assessments are time-consuming and costly to develop and validate properly. It is likely the primary reason qualitative research does not present more easily-accessible materials for instructors to implement in their classroom. Yet, these assessments are crucial as they provide an avenue to efficiently assess students' conceptual understanding of a mathematical topic. These research-based multiple-choice assessments provide the practical application of the RUME community to mathematics instructors. We feel that technology can aid researchers in transforming more open-ended questions and qualitative results into multiple-choice assessments that can be used by instructors. The next section will describe how we can dynamically generate quality multiple-choice items.

Automated Item Generation

We use a typical College Algebra item (Figure 1) to introduce multiple-choice item terminology. A *multiple-choice item* consists of a stem and options. The *stem* includes the context, content, and problem for the student to answer. In the example in Figure 1, this includes the instructions (context) and the problem. By *problem*, we refer to the content issue that must be solved. In the example in Figure 1, this would be solving the linear equation. Solving this problem leads to the *solution*. Plausible, but incorrect, answers to the problem are referred to as *distractors*. The solution and distractors are used to create the *options*, or choices presented that the student must choose from. Of these, the *correct option* corresponds to the option that correctly solves the problem in the stem (solution) while the *distractor options* correspond to the incorrect, distractor solutions.

There are currently two general strategies to generate distractors. The first strategy focuses on similarities between the solution and distractors. For example, a numeric solution could be manipulated in some form: being negated, divided by a factor, or shifted a small amount. Manipulating the solution in some way to make similar responses does not require a great deal of time and resources, and thus is commonly utilized (Gierl, Bulut, Guo, & Zhang, 2017). The disadvantage

[Stem]

Solve the linear equation below.

[Problem]

$$\frac{-3x-6}{3} - \frac{-8x-8}{5} = \frac{7x+6}{2}$$

[Options]

A. $x = -\frac{40}{29}$ **[Distractor]**

B. $x = -\frac{34}{29}$ **[Solution]**

C. $x = -\frac{66}{29}$ **[Distractor]**

D. $x = -\frac{17}{10}$ **[Distractor]**

Figure 1: Example of a typical multiple-choice item.

to this method is that distractors may not reflect actual student thinking. Students with incomplete knowledge may be able to eliminate these types of distractors and thus arrive at the solution (or, at least, more easily guess at the solution), thereby rendering the goal of assessing student knowledge moot. In short, multiple-choice items developed with these types of distractors would not provide feedback on students' potential cognitive processes.

The second method focuses on common misconceptions in student thinking while they reason about the problem. These misconceptions can be recalled and utilized by experienced content specialists reflecting on the common errors they have seen in the past or identified through evidence-based research on students' work during open-ended items (Gierl et al., 2017). This approach creates high-quality distractors that mirror responses students may make during an open-ended assessment.

In addition to the two methods above, we could consider how a student's conception of a mathematical topic would influence their response to the question. This strategy would enable the instructor to link certain multiple-choice responses to the student's conception at the time of the test. Carlson et al. (2010) utilize this method in the PCA to great success. By creating distractors based on a student's conception as it develops over time, instructors can more accurately assess and improve student understanding.

One avenue for creating quality distractors based on all three methods above is Automatic Item Generation (AIG). AIG utilizes computer technologies and content specialists (or evidence-based research) to automatically generate problems, solutions, and quality distractors. Few examples of AIG currently exist, even in the context of mathematics (Gierl et al., 2017; Gierl, Lai, Hogan, & Matovinovic, 2015). We will now illustrate how to leverage the knowledge of the research community to automatically generate distractors, and in doing so, generate ways to assess student knowledge as it develops.

Methodology

Dubinsky and Wilson (2013) investigated low-achieving high school students and their understanding of the concept of function. In their research assessment, they asked the following typical questions about composition of functions:

1. Suppose f and g are two functions. Find the compositions $f \circ g$ and $g \circ f$.
2. Suppose $h = f \circ g$ is the composition of two functions f and g. Given h and g, find f.
3. Suppose $h = f \circ g$ is the composition of two functions f and g. Given h and f, find g (Dubinsky & Wilson, 2013, p. 97).

Correct answers to these questions can provide some knowledge about students' understanding of functions in general. In fact, the authors state:

> In both the written instrument and the interviews, we asked students questions, some of which we considered to be difficult, about composition of functions. Our intention was to investigate the depth of their understanding of the function. We also felt that success in solving these problems was an indication of a process conception of function and in some cases, an indication of a process conception that was strong enough so that it could be reversed in the mind of a participant in order to solve a difficult composition problem (Dubinsky & Wilson, 2013, pgs. 96-97).

While open-response items would provide more information about students' understanding, this illustrates how correct answers to multiple-choice items could suggest students' conceptions of a particular concept. It is this belief that allows even multiple-choice questions to be used as learning tools in the classroom, as they can shed light on what students understand and allow instructors to challenge misconceptions. In order to be successful, multiple-choice items should include the common conceptions students may have. We illustrate how to develop quality distractors in the context of composition of functions below.

Consider the typical College Algebra exam item in Figure 2. The question requires students to compose two functions and evaluate the composition at a given point $x = a$. A student with adequate procedural understanding of function composition will compose the new function and evaluate it at the point to obtain $f(g(5)) = \left(\frac{1}{3}(5)^2 + 1\right)^2 = \frac{784}{9}$ which is answer choice A. in Fig. 2. Two other common responses Dubinsky and Wilson (2013) observed students made when solving function composition problems of this type were (a) composing the functions in an opposite order *(answer choice C.)* and (b) conflating the composition notation with multiplication notation *(answer choice B.)*. These responses would correspond to (a) a student recognizing composition as a new operation yet not performing the action correctly and (b) a student not recognizing composition as a new operation, similar to multiplication having multiple representations: $\times, \cdot$, and the absence of an explicit operator such as with $4x$.

To be clear - this question does not assess a student's conceptual understanding of composition of functions. It is however necessary students can illustrate adequate procedural knowledge of composition *before* moving on to develop a conceptual understanding of the operation. We now illustrate an automated question meant to assess a student's conceptual understanding of function composition.

Specific: Suppose $f(x) = (x+1)^2$ and $g(x) = \frac{1}{3}x^2$ are two functions. Find the composition $(f \circ g)(x)$ at the point $x = 5$.

A. $\frac{784}{9}$

B. 300

C. $\frac{1296}{3}$

Generalized: Suppose $f(x) = (x+c)^2$ and $g(x) = \frac{b_1}{b_2}x^2$ are two functions. Find the composition $(f \circ g)(x)$ at the point $x = a$.

A. $f(g(a)) = \left(\frac{b_1}{b_2}a + c\right)^2$

B. $(f \cdot g)(a) = \frac{b_1}{b_2}a^2 (a+c)^2$

C. $g(f(a)) = \frac{b_1}{b_2}(a+c)^4$

Figure 2: Typical College Algebra function composition exam item and generalized template.

The second and third types of function composition questions used by Dubinsky and Wilson (2013) required students to take a composed function $h(x) = (f \circ g)(x)$ and isolate the functions composing it. For example, given $h(x)$ and $g(x)$, a student would then be asked to find the expression for $f(x)$, or find the value of the expression at a given point a. While Dubinsky and Wilson (2013) did not provide alternative student responses, we constructed a question and two "potential" student responses in Figure 3. In this example, the student is given a table representation of the functions and asked to consider reversing the function composition to evaluate f at. Rather than reversing the composition, a student could evaluate $h(g(2))$ and "solve" the problem using similar steps to question 1. This would suggest the student has memorized a procedure to evaluate composition of functions, but does not recognize the need to reverse the process. Alternatively, a student could evaluate h at 2, then find the corresponding x value to when g is 1. This would suggest the student recognizes the need to reverse the composition process but the order of the function composition $f(g(x))$ was inverted. Finally, a student could state that without knowledge of the function f, they cannot evaluate any point. This would suggest the student views a function as a single algebraic formula.

The ability to reverse the composition and isolate the functions composed, as well as describe this process in general, would be growth of a conceptual understanding of composition. Now, a single multiple-choice question cannot provide an instructor with strong evidence of a student's procedural and/or conceptual understanding. However, by combining a series of questions linked to common conceptions on composition, instructors can identify where a student is in their conception and provide targeted feedback. With the added technological component, this identification can be

Specific: Given **only** the information in the following table, find $f(2)$ *(if possible)*.

x	$h(x)$	$g(x)$
2	1	-3
-3	4	1
-2	0	2

A. $f(2) = 4$ $[h(g(2))]$

B. $f(2) = 0$ $[g(x) = 2 \rightarrow h(x)]$

C. $f(2) = 1$ $[h(x) = 2 \rightarrow g(x)]$

D. It is not possible to find $f(2)$ based only on the information in the table.

General: Given **only** the information in the following table, find $f(a_1)$ *(if possible)*.

x	$h(x)$	$g(x)$
a_1	c_1	b_2
b_2	b_3	c_1
a_2	a_3	a_1

A. $f(a_1) = b_3$ $[h(g(a_1))]$

B. $f(a_1) = a_3$ $[g(x) = a_1 \rightarrow h(x)]$

C. $f(a_1) = c_1$ $[h(x) = a_1 \rightarrow g(x)]$

D. It is not possible to find $f(a_1)$ based only on the information in the table.

Figure 3: Example and template for function composition problems type 2 and 3.

automated and provided to the student without the instructor combing over the student's work. It is this fine-grained assessment and feedback that can improve how students develop their understanding throughout a course. In short, quality multiple-choice assessments can remove the time-burden of free-response assessments while (theoretically) providing similar results on student thinking.

Discussion

Generating multiple-choice assessments that can potentially indicate a student's level of understanding is attractive for a variety of reasons. From a practicality standpoint, these assessments would be cost-effective (both in time and resources to develop) and quick to grade. Providing the linked distractors for each item choice can also draw students' attention to their conception, allowing them to modify their thinking. It is in this way - explicitly challenging student conceptions in a cost and time effective manner - that these assessments can be used as practical learning tools. The mathematics education research community has the knowledge needed to create these quality multiple-choice assessments. By combining this knowledge with automatic item generation, the mathematics education research community can provide instructors with practical results based on empirical data.

In addition, theoretical frameworks such as APOS Theory posit learning trajectories students may take to learn a concept. By creating multiple-choice questions aligned to the various levels as students' knowledge develops, instructors can track a student's progress to provide individual feedback. With the ease that multiple-choice assignments can be graded, this individualized feedback can be scaled to large courses such as College Algebra and Calculus.

Automatically generated multiple-choice assessments can also serve as a research tool. We noted a paper by Dubinsky and Wilson (2013) in which they asked students to answer common college algebra questions. By converting these types of questions to multiple-choice items, researchers can widen their sample size to provide greater certainty of the results. Some authors in the RUME community have begun to tap the potential of multiple-choice assessments in research, such as Carlson et al. (2010) with their use of multiple-choice assessments in a precalculus concept inventory. A wide-spread use of multiple-choice assessments based on empirical evidence can provide the sample size needed to produce robust results.

References

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction*, *28*(2), 113–145. doi: 10.1080/07370001003676587

Dubinsky, E., & Wilson, R. (2013). High school students' understanding of the function concept. *The Journal of Mathematical Behavior*, *32*(1), 83–101. doi: 10.1016/j.jmathb.2012.12.001

Gierl, M., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082–1116. doi: 10.3102/0034654317726529

Gierl, M., Lai, H., Hogan, J., & Matovinovic, D. (2015). A method for generating test items that are aligned to the common core state standard. *Journal of Applied Testing Technology*, *16*, 1–18. Retrieved from `http://www.jattjournal.com/index.php/atp/article/view/80234`

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158. doi: 10.1119/1.2343497

Van Velzen, J. (2013). Educational researchers and practicality. *American Educational Research Journal*, *50*(4), 789–811. doi: 10.3102/0002831212468787

Vidakovic, D., Chen, G., & Miller, V. (2016). *NSF Grant 1624970: Promoting reasoning in undergraduate mathematics (PRIUM).*