Evolution of Test Items: From Open-ended to Multiple-Choice

Mary A. Sadowski
Purdue University

Sheryl A. Sorby
University of Cincinnati

Abstract

Grading is often a faculty member's least favorite chore, especially in engineering where open-ended problems prevail. For this reason, multiple-choice test items could be a popular alternative for assessing learning and understanding. In addition, most Learning Management Systems allow the instructor to create multiple-choice questions to be scored automatically by the system. The use of multiple-choice items in engineering graphics education could increase efficiency, allowing instructors to focus on other aspects of their teaching rather than spending significant time grading open-ended problems. The authors of the this paper have been involved in a project to develop a Concept Inventory for Engineering Graphics over the past several years. Since Concept Inventories typically consist of multiple-choice items, development of this instrument was reliant on the creation of numerous valid and reliable items. This paper will focus on the process employed in multiple-choice item creation with application to engineering graphics. The process will be illustrated through demonstration of item evolution through several iterations.

Background

According Educational Testing Services, (ETS) there are several key steps that they follow in the creation of a valid and reliable test (ETS, 2017), including:

1. **Define Objectives.** The type of questions to be considered in defining objectives for specific tests are: who will take the test? What skills or knowledge should be tested? What kinds of questions should be included? How long should the test be? What is the level of difficulty for the test?

2. **Form the Item Development Committee.** The responsibility of the committee is to: define the objectives of the test, ensure that test questions are unbiased, determine the test format, develop supplemental materials, write the test questions, and review the test questions.

3. **Writing and Reviewing Questions.** Each question on the test undergoes several iterations until the final format is established. The reviews ensure that there is only one correct answer and that it conforms to the desired test style.

4. **Pretest.** After the questions have been developed, they are pilot tested with a group of individuals who are similar in age and skill level to the desired final testing population. The results from this pilot testing will help to determine the difficulty of each question, any ambiguities in items, or

need for further revision or possible elimination all together. This step also examines distractors to determine if they need to be revised or replaced.

5. **Detecting and Removing Unfair Questions.** ETS then looks at the language as a whole used in the items to ensure that the language is appropriate and non-offensive. In this step, they examine performance by different groups with similar knowledge and skills to ensure items are unbiased and are measuring what they were intended to measure.

6. **Assembling the Test.** In this step, the test is reviewed by experts who were not involved in creating the test, to ensure that the answers are correct and any discrepancies are resolved.

7. **Making Sure that Test Questions are Functioning Properly. Further statistical analysis is** performed on the instrument after administration to ensure its reliability.

Although not all of these steps are applicable to the development of items for a graphics mid-term or final examination (for example, assembling the committee), consideration of these steps could lead to a better, more robust test that actually measures student learning. It should be noted, however, that this is essentially the process used by the Concept Inventory design team as they have worked over the past several years in the creation of the instrument. The team is now in the process of completing Step #7 on this list.

According to Malamed (2010), there are ten best practices for creating accurate multiple-choice test items. Although her rules were written specifically for elearning applications, many are relevant to questions that are being designed for an LMS system as well as for standard in-class administered examinations. The best practices identified by Malamed are:

1. **Items test comprehension and critical thinking, not just recall.** Many engineering faculty avoid multiple-choice testing because they feel that this type of test only assess rote learning and not critical thinking skills. Malamed advises to design items that go beyond this by asking students to interpret facts or evaluate situations.

2. **Use simple sentence structure and precise wording.** Faculty who create multiple-choice items should carefully examine wording to ensure that there is no ambiguity or colloquial expressions that may not translate across cultural groups.

3. **Place most of the words in the question's stem.** This best practice ensure that the answer options are short and not confusing.

4. **Make all distractors plausible.** Wrong answers should be completely reasonable with no "give-away" distractors that hinder your ability to discriminate among test-takers.

5. **Keep all answer choices the same length.** Test-takers might be able to guess the correct answer merely by looking at the length of each choice.

6. **Avoid double-negatives.** Questions that include double-negatives are often confusing to the test-taker. For example, "Which of the following comments would not be unwelcome in a work situation?" could be replaced with "Which of the following comments are acceptable in a work situation?"

7. **Mix up the order of the correct answers.** You should make sure correct answers are randomly located and do not form a pattern.

8. **Keep the number of options consistent.** Typically, each question will have 3, 4, or 5 options. Test creators should decide how many options they will use throughout the test and then stick with it.

9. **Avoid tricking test-takers.** The objective of the test should be to assess learning. Therefore, options that can be interpreted in more than one way or that are too similar to one another should be avoided.

10. **Use "All of the Above" and "None of the Above" with caution.** The option of "All of the Above" encourages guessing and the option of "None of the Above" doesn't really assess what a student knows so these options should be avoided when possible.

Not all of these best practices may be relevant to the creation of multiple-choice items that assess graphics learning; however, test designers may want to keep these in mind as they go about designing accurate tests.

Developing Items for the Graphics Concept Inventory

The technique used to create the Concept Inventory was based on the Assessment Triangle as outlined by Streveler et al (2011); however, the technique described by ETS closely mirrors the process used for individual item development. In the first step outlined by ETS (ETS, 2017), a rigorous process for defining the topics to be covered by the test is advocated. In the case of the Graphics Concept Inventory, a previous project conducted a Delphi study to identify the topics for inclusion on the resulting instrument (Sadowski & Sorby, 2014; Sadowski & Sorby 2015). In the second development step, a team of faculty was assembled with a combined experience of more than 50 years in graphics education. In step 3 of the ETS process, the development team constructed more than 60 open-ended graphics questions. Questions were created to cover all of the topics identified in the Delphi study and were based on the experiences of the design team. A sample open-ended item from this step is shown in Figure 1.
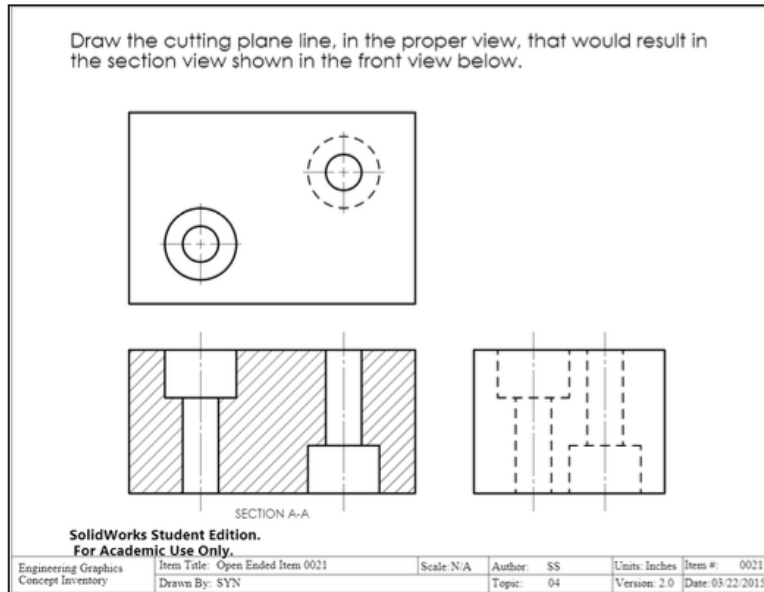
Figure 1. Sample open-ended item from Graphics Concept Inventory.

The item shown in Figure 1 was designed to test conceptual understanding of Offset Section views. In the next step of the Concept Inventory creation, the open-ended items were pilot-tested with students enrolled in graphics courses at three different institutions. The purpose of the pilot-testing was twofold—to determine relative difficulty of the items as well as to develop a battery of student-generated incorrect answers that could eventually be used for distractor creation. Figure 2 shows an incorrect student response to the question from Figure 1.
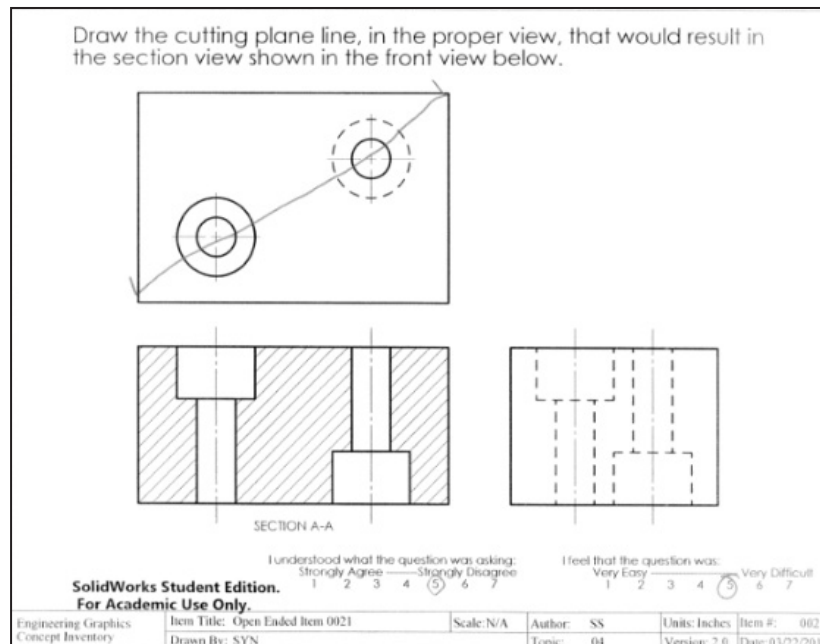

Figure 2. Sample student incorrect response.

For the next step in the Concept Inventory development, distractors were developed that were based on the most popular incorrect student responses to the open-ended problems. Based on the procedure outlined by ETS, steps 5 and 6 were iteratively completed in order to ensure that the test items were behaving appropriately (not too easy and not too hard), were clear and unambiguous (no more than one correct response per item; clearly worded) and there was consistency across items (4 choices per item). The final item for the problem from Figures 1 & 2 is shown in Figure 3.
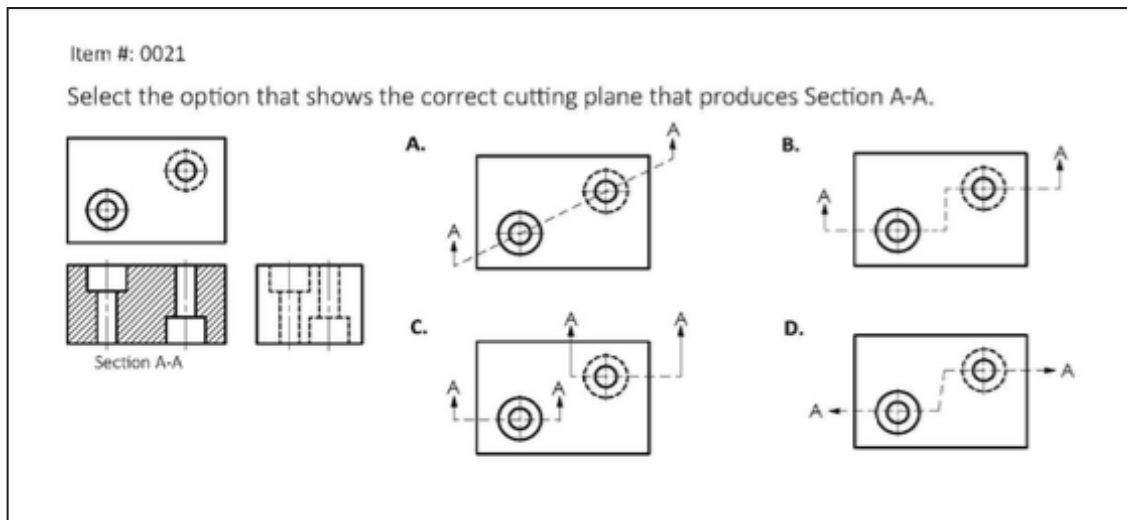

Figure 3. Final Concept Inventory item for the topic of Offset Sections.

The development team is currently completing step 7 from the ETS process by performing a final implementation and statistical factor analysis of data. Results from this final step will be forthcoming.

In the creation of the graphics Concept Inventory, many of the best practices listed by Malamed (2010) were employed. For example, all items on the test include 4 choices with one and only one correct response and three incorrect responses. Items test the ability of students to apply knowledge of graphics to novel problems—no definitions are queried. Distractors were developed from actual mistakes made by students so they likely meet the "plausibility" criteria and none of the questions uses either "All of the Above" or "None of the Above." The test design team labored over stem wording to ensure clarity and precision and since these are graphics items, most of the words are in the stem of the questions.

Conclusions

The need for a Concept Inventory in Engineering Graphics is well-established (Nozaki et al, 2016) and the use of Concept Inventories has been shown to lead to curriculum

innovation in a variety of disciplines (Evans et al, 2003). The Graphics Concept Inventory development team employed best practices from multiple sources in the creation of the items for this instrument. Final statistical factor analysis and hosting options for the instrument are in progress. The process the team members used could be applied to the creation of test items in graphics education as faculty attempt to streamline their grading without compromising the quality of their assessment.

Acknowledgement

References

Educational Testing Services (2017). *How tests and test questions are developed*. http://www.ets.org/understanding_testing/test_development/.

Evans, D. L., Gray, G.L., Krause, S., Martin, J., Midkiff, C., Notaros, B. M., Pavelich, M., Rancour, D., and Reed-Rhodes, T. Progress on concept inventory tools. In *33rd ASEE/IEEE Frontiers in Education Conference Proceedings*. (Nov. 2003), Boulder, CO pp. T4G1 – T4G8.

Malamed, C. (2010). 10 Rules for writing multiple choice questions. http://theelearningcoach.com/elearning_desig 2010n/rules-for-multiple-choice-questions/

Nozaki, S., & Study, N. E., & Steinhauer, H. M., & Sorby, S. A., & Sadowski, M. A., & Miller, R. L. (2016, June), *"Development of a Concept Inventory for Engineering Design Graphics."* Paper presented at 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana. 10.18260/p.26794

Sadowski, M. A., & Sorby, S. A., "Defining Concepts for an Engineering Graphics Concept Inventory: A Delphi Study," Proceedings of the 69th Midyear Conference of the Engineering Design Graphics Division of ASEE, Normal, IL, October 2014, CD-ROM.

Sadowski, M., & Sorby, S., "Engineering Graphics Concepts: A Delphi Study," Proceedings of the Annual Conference of ASEE, Seattle, WA, June 2015, CD-ROM.

Streveler, R. A., Miller, R.L., Santiago Roman, A.I., Nelson, M.A., Geist, M.R., & Olds, B.M. (2011). Using the "assessment triangle" as a framework for developing concept inventories: A case study using the thermal and transport concept inventory. *International Journal of Engineering Education*, 27(5), 1-17.

About the Authors

The authors can be contacted at—**Mary Sadowski**, Email: msadowsk@purdue.edu and **Sheryl Sorby**, Email: sheryl@mtu.edu.