The Effect of Item Order on Student Scores

Dennis Christopher, MS, Assessment Specialist
Questar Assessment Inc.

What is the most effective way to set up items on a test for a test blueprint?

If a student asks his teacher what fuel is used in the external tank of the space shuttle and then follows that up with a question about the requirements to become an astronaut, we would likely say those questions are random. For the first question we would have to call up information, also known as building a mental set, about the types of fuel and where they are housed in the full stack of the shuttle. With the next question, we would have to switch mental gears to recall the type of training required to be an astronaut. In the context of how we live, this scenario represents neither the way we think nor the way information is taught or assessed in a classroom.

According to Britannica.com, "[a] 'mental set,' or 'entrenchment,' is a frame of mind involving a model that represents a problem, a problem context, or a procedure for problem solving." ("mental set | psychology | Britannica.com," n.d.) A person being asked a question or challenged to solve a problem must go through the time-consuming task of mentally pulling together all the information she knows about a topic. In a typical conversation, the topics being discussed are connected to each other, allowing participants to more easily access their stored memories and respond more quickly. We use the term non sequitur to describe statements or questions that do not connect in a clear or logical way to anything said before it.

In the classroom, a student receives a set of information. She applies that information to some type of exercise. Then she must answer a series of questions or solve a series of problems to assess her understanding of the topic. While the student learns this unit, she may be asked to write a short essay, design and build a graphic, or answer a set of specific questions. All of these activities revolve around a single set of information.

Not only are students taught a single set of information at a time, they are also taught each set of information in a specific order. In education, we call that ordering "scaffolding". We build students' knowledge of a topic by starting with the most general information and moving to the most specific. We don't teach a child how to read by handing him a novel. We start with the building blocks of letters and letter combinations and work up to words and sentences. Learning is like putting together a puzzle — mentally working out how different pieces of information connect to each other to form a complete and coherent picture.

Students learn by connecting coherent information. Doesn't it make sense that connected information would be more easily and quickly retrieved?

> In the "Atlas of Science Literacy" (2001), Harold Pratt wrote: "To achieve coherence, a curriculum must build new ideas and skills on earlier ones — within lessons, from lesson to lesson, from unit to unit, and from year to year. With coherence in the curriculum the concepts and processes that students learn and become more complex as they construct new ideas and develop new skills. Research has demonstrated that what students already know about a topic is one of the most significant factors in determining their success in learning new, related content."

Given that teaching and learning typically focus on a specific unit of information, with sub-units related to that unit of information introduced in a specific order — and that most activities outside of the classroom, like becoming goalie for the soccer team or learning to play the drums, are similarly structured — why do most exam questions jump from topic to topic, with no logical structure or order? For example, the first question on a science exam is about DNA structure and the second question is about stomach acidity. Although a serious debater could make a case that these two questions are connected, the rest of us see two random pieces of information with no obvious connection to each other. Ordering the two questions this way forces the student to change his mental set to answer each question.

Considering the way humans learn, would students do better on exams written with items that follow the curriculum strands?

Limited research has been conducted on how item order affects test results, with a focus so far on item presentation in sequence, random order, or reverse sequence as taught in the order of the classroom curriculum. William Balch in "Item Order Affects Performance on Multiple-Choice Exams" (Balch, 1989) found that students who took an exam consisting of sequential items scored three percentage points higher than students who took an exam of randomly ordered items. The sequential exam students also scored four percentage points higher than students who took an exam with items that were listed randomly within a block of items that centered on a single chapter.

Balch suggests that sequentially ordered exams provide retrieval cues, which may help with memory recall that is consistent with encoding specificity, or how information is connected within the brain. The context of surrounding information used in encoding is utilized in information retrieval, and the sequential test question order provides a situation where context of encoding and retrieval is similar. In addition, Balch found that there was no significant difference in completion times between these versions of the exam.

> Balch concluded: "The results suggest that exam scores can be influenced by item order. Students scored slightly higher with the sequence form (S) than with the CC (random inside of a chapter) or the R (totally random) forms. Such differences in scores are consequential in many course grading systems."

This research would indicate that item progression can be a contributing factor to exam scores. Many researchers and psychometricians believe that the ordering of items may lead to cueing. In other words, it might be possible to "lead" the student from one correct answer to another. If that is the case, it would be incumbent on the test developer and the psychometricians to make sure that one item would not give away, or cue, the answer to another question.

Other researchers have found that the order of items on a test resulted in no significant difference in scores — except the scores of a select group of students identified by the study as "students who experience intense test anxiety." In this study conducted by Pettijohn and Sacco (2007), students were asked about their test anxiety when given sequential (S) and random order (RA) item exams.

> Pettijohn and Sacco concluded: "Our post-exam anxiety assessments showed that participants reported greater anxiety in response to the RA (random order) test question

> order condition compared to the S (sequential) condition.... Overly anxious students, students with learning disabilities, or students with other special circumstances may benefit from sequential test question order."

Although research has produced conflicting conclusions over the scores produced by random versus sequentially ordered items, stress appears to be reduced for students who take sequentially ordered exams. As discussed in the Questar Assessment white paper "Performance and Assessment" (Christopher and Sewall, 2015), critics of standardized testing often express concern about the stress these tests place on students. "Performance and Assessment" outlines ways to reduce that stress both before and during the exam. Sequentially ordered items on exams, along with teaching students memory, retention, and retrieval skills, may reduce stress on students and increase test scores.

References

mental set | psychology | Britannica.com. (n.d.). Retrieved from http://www.britannica.com/EBchecked/topic/1442593/mental-set

Project 2061 (American Association for the Advancement of Science). (2001). Atlas of science literacy. Washington, DC: American Association for the Advancement of Science.

Balch, W. (1989). Item Order Affects Performance on Multiple-Choice Exams. Teaching of Psychology, 16(2), 75-77. doi:10.1207/s15328023top1602_9

Project 2061 (American Association for the Advancement of Science). (2001). Atlas of science literacy. Washington, DC: American Association for the Advancement of Science.

Christopher, D., & Sewall, L. (2015, June 4). Using Performance Assessments to Bridge the Gap between Accountability and Learning. Retrieved from http://www.questarai.com/wp-content/uploads/sites/2/2015/06/Using-Performance-Assessments-to-Bridge-the-Gap-between-Accountability-and-Learning.pdf