Automating the Generation of Test Items

2

Hollis Lai
University of Alberta, Canada

Mark Gierl
University of Alberta, Canada

INTRODUCTION

With an ever increasing demand for more knowledge in our economy, organizations require methods of assessment to evaluate the skillset of their workers to produce new ideas, make new products, and provide new services. The ability to create these ideas, products, and services will be determined by the effectiveness of our educational programs. Education provide students with the knowledge and skills required to think, reason, communicate, and collaborate in a world that is shaped by knowledge services, information, and communication technologies (e.g., Binkley, Erstad, Herman, Raizen, Ripley, Miller-Ricci, & Rumble, 2012; Darling-Hammond, 2014). Educational testing has an important role to play in helping students acquire these foundational skills. Educational tests, once developed almost exclusively as a right-of-passage to satisfy demands for accountability and outcomes-based summative testing, are now expected to provide teachers and students with timely, detailed, formative feedback to directly support teaching and learning. With an increasing focus on formative learning principles being adopted to guide our educational testing practices, where assessment-related activities provide constant and specific feedback to modify and improve learning, assessments are being administered more frequently (Black & Wiliam, 1998, 2010). But when testing occurs more frequently, more test items are required. These additional test items must be created efficiently and economically while maintaining a high standard of quality. Fortunately, this requirement for frequent and timely educational testing coincides with the dramatic changes occurring in educational technology. Developers of local, national, and international educational tests are now implementing computerized tests at an extraordinary rate (Beller, 2013). Computerized testing offers many important benefits to support and promote key principles in formative assessment. Computers permit testing on-demand thereby allowing students to take the test at any time during instruction; items on computerized tests are scored immediately thereby providing students with instant feedback; computerized tests permit continuous administration thereby allowing students to have more choices about when they write their exams. In short, developments in computing technology enables the infusion of formative principles into our testing practices that would not have been previously possible.

Despite these important benefits, the advent of computerized testing has also raised formidable challenges, particularly in the area of test item development. Tests now require access to large numbers of diverse, high-quality test items to implement computerized testing as items are continuously administered to students. Hundreds of items are needed to develop the test item banks necessary for computerized testing. Unfortunately, creating test items is a time consuming and expensive process. Each individual item is written, initially, by a content specialist and, then, reviewed, edited, and revised by groups of content specialists (Gierl & Lai, 2016; Rudner, 2010). Hence, item development has been identified as one of the most important problems that must be solved before we can fully migrate to computerized

DOI: 10.4018/978-1-7998-3473-1.ch019

testing because large numbers of high-quality, content-specific, test items are required (Haladyna & Rodriguez, 2013; Webb, Gibson, & Forkosh-Baruch, 2013).

One emerging test item development method used to address this challenge is *automatic item generation* (AIG) (Gierl & Haladyna, 2013; Irvine & Kyllonen, 2002). AIG is a relatively new but rapidly evolving research area where cognitive and psychometric modeling practices guide the production of tests that include items generated with the aid of computer technology. Research on AIG has adopted different perspectives, including the use of natural language processing and artificial intelligence (e.g., Gütl, Lankmayr, Weinhofer, & Höfler, 2011; Moser, Gütl, & Lui, 2012; von Davier, 2018), frame-semantic representations (e.g., Cubric & Toasic, 2010; Higgins, Futagi, & Deane, 2005), schema theory (e.g., Singley & Bennett, 2002), and sematic web-rule language (Zoumpatianos, Papasalouros, & Kotis, 2011; Leo et al., 2019). The purpose of this chapter is to describe and illustrate the most practical method for generating test items, which is template based. By template-based AIG, we mean methods that draw on *item models* to guide the generative process. Gierl and Lai (2013, 2016, 2018) developed a three-step process for template-based AIG. In step 1, content specialists create a cognitive model for AIG. A cognitive model is a representation that highlights the knowledge, skills, and content required to generate new test items. In step 2, an item model is developed to specify where the cognitive model content is placed in each generated item. An item model is a template that highlights the variables in a test item that can be manipulated to produce new items. In step 3, computer algorithms place the cognitive content into the item model. With this process, hundreds of items can be created from a single item model. The purpose of this chapter is to describe how AIG can be used to generate test items using the selected-response (i.e., multiple-choice) format. To ensure our description is both concrete and practical, we illustrate template-based item generation using an example from the complex problem-solving domain of the medical health sciences. The chapter is concluded by describing two directions for future research.

BACKGROUND

Gierl and Lai (2013, 2016, 2018) described a three-step approach for template-based AIG. In step 1, a content specialist creates a cognitive model for AIG. In step 2, an item model is developed to specify where the cognitive model content is placed in each generated item. In step 3, algorithms place the cognitive content into the item model.

Step 1: Identify Content For Item Generation

To begin, the content for item generation is identified by the content specialists. This content is identified using design principles and guidelines that highlight the knowledge, skills, and abilities required to solve problems and perform tasks in a specific domain. A cognitive model for AIG is a representation that organizes the cognitive- and content-specific information into a structured representation of how the content specialist expects that examinees will think about and solve problems. Recently, Gierl and Lai (2016) proposed the *key features* cognitive model for AIG. With this model, item generation is guided by the relationships among the key features specified in the cognitive model. It is used when the attributes or features of a task are systematically combined to produce meaningful outcomes across the item feature set. The use of constraint programming in step 3 of the AIG process ensures that the relationships among the features yield meaningful items. The key features cognitive model is most suitable