






Future of Assessment: Automatic Item Generation

Bharati B. Belwalkar, PhD | Christina Curnow, PhD

Acknowledgment: Authors thank Carol Wong (Industrial/Organizational intern) for her assistance in conducting the literature review for this research note.

In the ever-evolving landscape of education and employment testing, the assessment industry is on the brink of a revolutionary transformation. Automatic Item Generation (AIG) is set to disrupt traditional assessment methods, empowering educators and assessment-developers with the potential to swiftly create and adapt high-quality test items. In this research brief, we discuss a seismic shift in the way we assess knowledge and skills! We provide an overview of AIG along with the pros and cons of its two approaches: template-based and non-template AIG.

The traditional approach to item development is a time-consuming and expensive process, because each individual item is written by subject matter experts (SMEs). The rise in popularity of computer-based testing (CBT) contributes to an increased need for a large number of diverse and high-quality test items.¹ As traditional item development cannot effectively sustain this demand, AIG serves as a potential solution to promote more efficient item writing. Compared to traditional item development, AIG has a number of practical benefits:

-  **More time- and cost-effective** as a single model can generate a large number of assessment items
-  **Potential to reduce human involvement** in the item development process with structured guidelines and/or algorithms, which may help reduce subjectivity
-  **More flexibility in creating updated items**, which is particularly useful in domains that are constantly changing (e.g., health science)
-  **Decreased item exposure** because of the sheer size of the AIG test bank resulting in enhanced test security
-  **Easier and more efficient monitoring** of item quality and **reduced common errors** in the review process for items generated using a template

Conceptually, approaches to AIG can fall into one of the two general categories: template-based AIG and non-template AIG. Template-based AIG refers to methods that use templates generated by SMEs to guide item generation.² Whereas non-template AIG aims to minimize human involvement in the item development processes to create the essential features of an item model. This modeling approach often relies on natural language processing (NLP) techniques to generate item features based on a

corpus of data that are directly related to the content domain where item generation is intended to occur. Exhibit 1 summarizes key advantages and limitations of both AIG approaches.

Exhibit 1: Key Advantages and Limitations of Template-Based and Non-template AIG approaches

Template-based AIG	Non-template AIG
Advantages	
<ul style="list-style-type: none"> • More practical as simple item models can be developed with relative ease • Helpful in ensuring that the generated items address the intended purpose of the test 	<ul style="list-style-type: none"> • Minimized human involvement • Generating items directly as no template is required • Used for generating gap-fill or Wh-type questions
Limitations	
<ul style="list-style-type: none"> • Heavy reliance on subject matter expert (SME) involvement • Vulnerability to test coaching due to similarity among items • Challenging in domains where explicitly extracting features and values to represent content area (e.g., reading comprehension) 	<ul style="list-style-type: none"> • Challenge of obtaining a corpus of data that are directly related to the test purpose • High proportion of “unacceptable” items or items that require human revision

In theory, AIG could solve several practical item-development issues by offering increased larger pools of items with varying complexity, fewer item exposure concerns, flexibility of test administration, limited human reviews, and reduced resource needs. However, operationalizing AIG is a resource-intensive endeavor, requiring consideration of multiple factors—volume of testing, item types, technology, and feasibility of piloting AIG content, to name a few.

To our dismay, very little guidance exists in the current AIG literature for the testing programs that want to decide if AIG is right for them. We guide our clients by exploring specific factors to consider when deciding to explore AIG and teach them to consider the critical elements that decide whether to adopt, delay, or discard the AIG idea.

Reach out to us if you are considering AIG but do not know where to start. We, at the [Human Capital Solutions](#) hub at AIR, would be happy to help!

References

- ¹ Casillas, A., Dages, K. & Ferrell, B. (2019). Challenges, trends, and opportunities of the testing industry: Practitioners' perspectives. *Industrial-Organizational Psychologist*, 56(4).
- ² Gierl, M. J., & Haladyna, T. M. (Eds.). (2012). *Automatic item generation: Theory and practice*. Routledge.