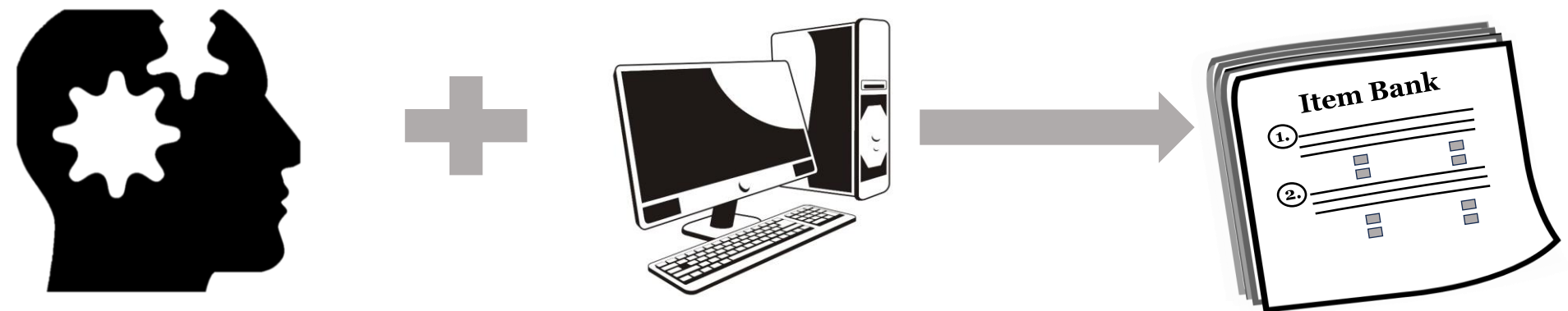


What is Automatic Item Generation?

- High-quality items are required for educational testing in order to accurately measure student knowledge, skills, and competencies. With the rapid growth in computer-based testing, students can write tests more frequently. But this also means that instructors need a large bank of test items to ensure that students receive a continuous supply of new content and to ensure that items in the bank are not over-exposed.
 - Item development (as we all know) is a tedious process. The traditional method for item development—where we write one item at a time—is time-consuming. It is also expensive. *Just one operational item* for a high stakes exam can cost between \$1500-\$2000. The traditional method is simply becoming insufficient in keeping up with modern testing and there must be a faster way to implement this process:
- 
- Automatic Item Generation (AIG) was developed as a solution to address the rising need for large item banks required for computer-based testing. AIG is an augmented intelligence approach for creating item banks. AIG combines subject-matter expertise with modern computing to produce a large numbers of items efficiently and economically.

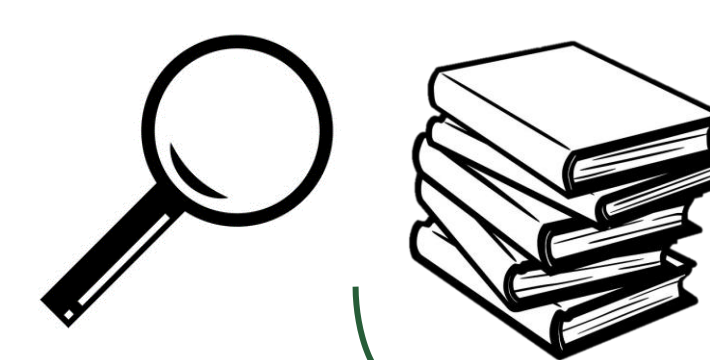
Item Definition

- Although there are multiple definitions of the word “item” in the literature most, if not all, definitions encounter the “black box” problem. The definitions describe the input and output but offer no explanation on how to transform the input into the output.
- To address this problem, Gierl, Lai, and Tanygin (2021) defined an item as “an explicit set of properties that include the parameters, constraints, and instructions used to elicit a response from the examinee.”
- By defining the input as parameters, constraints, and instructions, we are able to explicitly define the requirements of item development thereby allowing for replication.

Three Steps for Item Development

- Gierl, Lai, and Tanygin (2021) developed a three-step method to implement AIG. This method requires an explicit description of the development of each test item. This is counter to the traditional approach that depends on the subject-matter expert’s experience and creativity to develop an item from scratch. The method follows these steps:

- The subject-matter expert creates a cognitive model by identifying the content that will be used to produce new items.
- The subject-matter expert creates an item model by extracting the content from the cognitive model and explicitly identifying the features that can be manipulated to create new items.
- A computer uses the item model to generate a large number of new items.



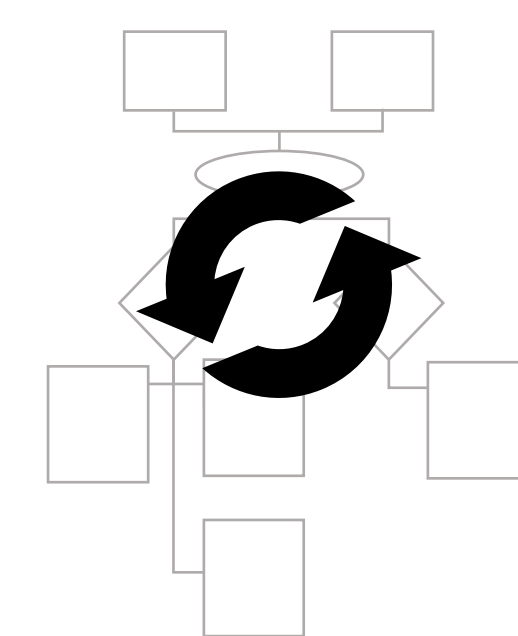
Item Model:	
Stem	A 14-year-old female sees her doctor and reports that she's been experiencing (Bony Aches) that have developed (Onset). What is the most likely diagnosis?
Element	Age: 10-16, Sex: F Body Aches: 1. slight aches, 2. slight pains, 3. severe aches, 4. severe pains Onset: 1. over a few days, 2. within 3-6 hours, 3. suddenly
Key	Common cold; Seasonal flu



Template-Based AIG

- Template-based AIG requires heavy investment from the subject-matter expert to develop the cognitive and item models that a computer, in turn, can use to generate items. There are important benefits when using template-based AIG, including:

- The subject-matter expert must define the item's purpose, which clarifies the item creation process.
- A robust model has the capability to produce hundreds or thousands of unique items.
- Problematic items can be quickly fixed by updating the cognitive model.



GPT-Based AIG

- Prior to the rise in large language models, AIG depended on specialized software to integrate SME instructions with computational capabilities. With the advancements of large language models, specifically OpenAI’s GPT models, there is no longer a need for specialized software. GPT may be a more accessible solution to generating items and thus facilitating a novel approach to AIG.
- Armoush and Gierl (2024) demonstrated GPT’s ability to quickly generate items across many content areas. Yet, its ease of use can lead to subpar items that are unusable in assessments. To counter this problem, we have found that providing clear, specific instructions is imperative for generating high-quality items.
- Sayin and Gierl (2024) created a template-based automatic item generation method using GPT-3.5. They used a cognitive model to provide instructions to GPT for generate reading comprehension items. Selected items were evaluated by subject-matter experts and were found acceptable or required only minor revisions. The generated items also had strong statistical properties when piloted with students.

Future Work

- As large language models advance, they will continue to become easier to use for item generation. However, there is a clear link between the quality of the items and the investment spent on their creation. While advancements are expected to support the capabilities of subject-matter experts, they are not sufficient to replace the essence of human creativity and skill required for developing high-quality educational test items.

References

- Armoush, N., & Gierl, M. (2024, April). *Using large language models for automatic item generation to create multiple-choice questions*. Poster presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Gierl, M. J., Lai, H., & Tanygin, V. (2021). *Advanced methods in automatic item generation*. Routledge.
- Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to Generate Reading Comprehension Items. *Educational Measurement: Issues and Practice*.