Assessment Engineering Summary Report

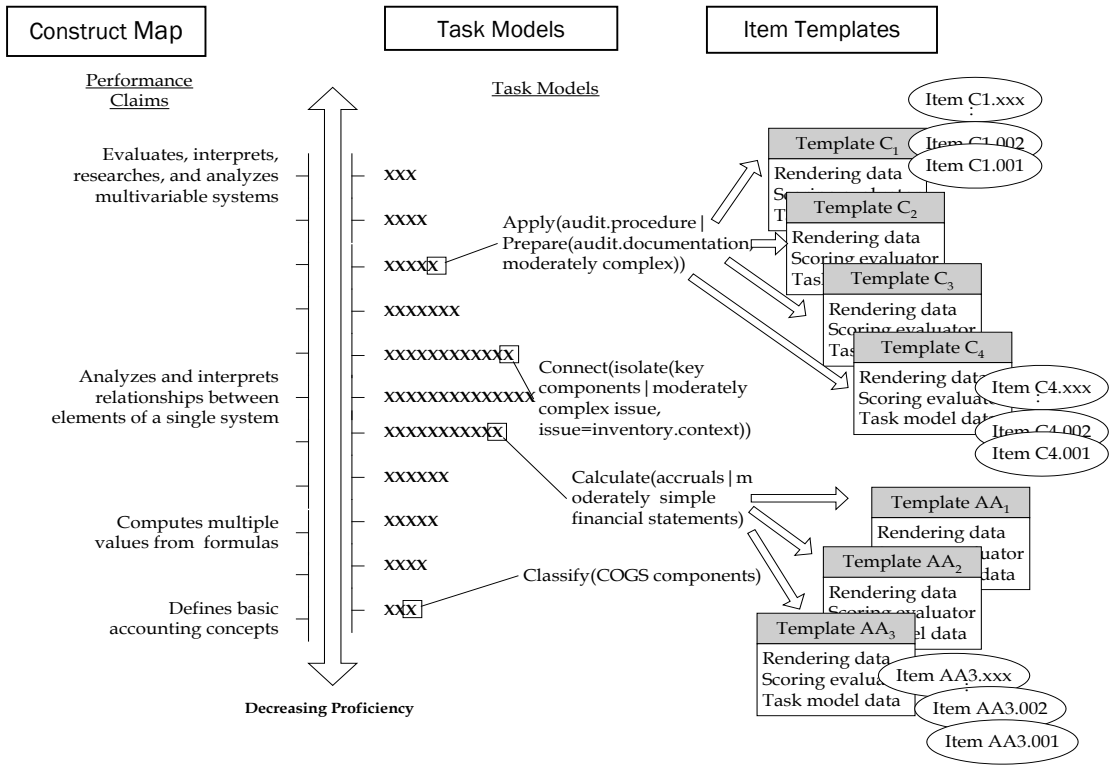
Presenter: Matthew Burke, PhD
Director of Exam Development
National Commission on Certification of Physician Assistants

Background

Assessment Engineering is one of a class of principled assessment frameworks that provides the potential for better test development using a confirmatory approach. It is a new way to design and implement scalable, sustainable, and ideally lower-cost examinations by combining cognitive modeling with engineering principles. It can offer improvements in construct and response processing validity, increase item production, and enhance score reporting. Assessment engineering necessitates a fundamental change in how one thinks about test development as one now relies on a cognitive model rather than content blueprints to develop items. These models are also used for scoring and to provide more formative feedback to the candidate.

Key Points from Presentation

Dr. Matthew Burke began with an overview of the three components of Assessment Engineering: Construct Map, Task Models and Templates as shown in the below figure:



The construct map is the big picture view. It allows a visual representation of the score scale and demarcates levels of proficiency relative to the scale (from low levels of proficiency – e.g., recall items of basic content knowledge– to high levels of proficiency – e.g., items that require evaluating, analyzing or interpreting complex content knowledge). Task models are families of items which are all at comparable levels on the construct map. The most detailed level is the Template. Templates provide specific instructions to generate items that fit each task model (blueprint). Current items can be reverse engineered and used to create Task Models, so that the blueprint is clearly defined with respect to knowledge demands and level of complexity. This is then used to help item writers develop questions targeting specific concepts and levels of thinking. Assessment engineering also allows for automated item generation (AIG), which has the potential to greatly increase our item banks. The main challenge of assessment engineering is that there is no "off the shelf" product that creates Construct Maps, Task Models, or Templates. The ABP would have to take their current blueprints and reverse engineer them, which requires extensive up-front work. However, the work is an investment which pays off later with ease of item writing and the potential for AIG.

Key Points from Breakout Session

Potential ways that assessment engineering could be used are: item writing and generation, enhancing feedback to diplomates, and clearly delineating knowledge expectations. Using a defined framework, it is very clear to item writers what types of questions they are to write and how complex they should be. Under the current system, it takes about 2 to 3 years for an item writer to be able to write good test questions. Assessment engineering may significantly reduce the time it takes to become a proficient item writer. Assessment engineering also lends itself to AIG, which will increase the item pool at a faster pace than current practices will. Because of the direct link between the knowledge and skills specified in the task models and the proficiency claims of the construct map, feedback could be more robust and help examinees identify weaknesses and address knowledge gaps. Knowing the complexity of items, and the components that drive the complexity of items, also allows the item writers to have more control over the complexity of the test, so that the ABP could produce test forms targeting different ability levels (e.g., a multistage adaptive test). The ABP must continually revisit and update their constructs/blueprints, which would keep content current and relative to the rapidly changing healthcare world. A disadvantage of assessment engineering is that it does require work up-front to reverse engineer current items. Time is needed to alter how we define our test content.

Conclusions

Assessment engineering may allow the ABP to build a better exam. It may help with item generation and enable the ABP to provide enhanced feedback to examinees. The ABP has staff that is knowledgeable with the framework and experienced in its practical applications, so the ABP is uniquely prepared among the ABMS member boards to begin exploration of the capabilities of this framework. Pursuing assessment engineering will make the potential testing innovations that rely on a robust item pool more plausible. Providing targeted, formative feedback to diplomates based on exam performance will also promote learning and physician improvement. Finally, outside of the time spent on development, assessment engineering is a very low risk innovation. If the ABP pursues assessment engineering and decides that the costs outweigh the benefits, the ABP will have a better understanding of the construct, item models/templates for future item development, and items that can be pretested.