EMIGS Cognitive Examination Development Process

The development of the EMIGS Cognitive exam began with a formal Job Task Analysis (JTA) process that was conducted during a two-day in-person meeting. The meeting was facilitated by a psychometrician and was attended by eleven Subject Matter Experts (SMEs). The SMEs were carefully selected academic faculty members from top university programs in the US and Canada, with expertise in resident education and assessment. The first outcome of this meeting was to establish the Test Definition Document (TDD), which outlines the purpose of the exam, the target audience, and the required minimal competency required to pass the exam. The second outcome was to identify all of the knowledge, skills, abilities, and judgements (KSAJs) that are relevant to the intended audience and ability level. Finally, the group generated specific testing objectives from the KSAJs to clearly indicate what a minimally competent candidate is expected to be able to demonstrate and what evidence of this ability is considered acceptable. These testing objectives, along with the definition of the minimally competent candidate, provided very explicit guidance to the item writers, the subsequent qualitative reviews, and approvals described later.

In order to ensure that the number of items per objective was appropriate for the relative importance of each objective, a blueprint survey was generated. The survey was distributed to nearly 100 SMEs to have them rate 1) the importance, 2) the relevance, and 3) the frequency for every objective on the exam. The responses were analyzed to determine the exact number of questions that should be on the exam for each objective; this is referred to as the exam blueprint.

Test questions are generically referred to as test items since some are technically not asked in the form of a question. Item development for the cognitive exam began by a psychometrician training the SMEs on item writing . The SMEs were trained on the latest research in how to construct the test items to maximize their validity and reliability. Specific attention was given to ensuring that items were in complete compliance with the intended testing objectives and that the items were written at the right difficulty level; meaning that minimally competent candidates should be able to answer the questions correctly. The SMEs were also taught other best practices to maximize the future statistical performance of the items.

After items were completed by a subject matter expert, a psychometrician edited each item to verify that all of the psychometric rules had been followed, that a reference was noted for each item, and that a rationale had been provided for every wrong answer. A language review was then performed to check for grammar, readability, misleading or confusing wording, and sensitivity to cultural, racial, and gender biases.

Technical review meetings were held with six to ten SMEs in attendance and were each facilitated by a psychometrician. The group of SMEs revised each item until the entire

May 2023

group came to a consensus that 1) the item was congruent to the objective, 2) the item was written at the right difficulty level, 3) the correct answer(s) was/were always correct, 4) the wrong answers were specious, and 5) the items were relevant and important. Any item where the group was unable to reach a consensus was either sent back to the item writer with instructions to modify it, or the item was rejected if modification was not feasible.

All items were then published on an alpha test and administered to 54 candidates and a full statistical analysis was conducted. Poorly performing items were eliminated, and marginal items were identified and either reworked or replaced. A beta test was then administered to 52 additional candidates and another full statistical analysis was conducted. Marginally performing items were again eliminated and the remaining eligible items were used to build the equated final forms of the exam. The final forms were built to the blueprint specifications and were balanced in difficulty, discrimination, reliability, and time. The form reliability statistic (KR20) for the forms was 0.92, indicating a very strong reliability.

With the forms now built and equated, a standard setting panel was convened to determine the appropriate cut score for the exam. A group of SMEs met in an on-line meeting to review the entire exam and to determine what score minimally competent candidates would be expected to achieve on the exam. Each SME that participated in this study was a medical school faculty member and was intimately familiar with the desired ability level of the target audience for the exam. The beta participants were then scored based on the final forms and the cut score to determine their pass/fail status. The final forms were again independently reviewed by the co-chairs of the EMIGS Steering Committee before releasing the exam.

Since the final forms of the exam have been published and administered to the target audience (OBGYN Residents), the pass rate has increased as anticipated, likely secondary to the candidates preparing to take the exam.

May 2023