Are Simulations Effective as Job Performance Assessment Tools?

Sowmya Ramachandran
Stottler Henke Associates
San Mateo, CA
USA
sowmya@stottlerhenke.com

Erik Sincoff
Stottler Henke Associates
San Mateo, CA
USA
esincoff@stottlerhenke.com

Marissa Shuffler
Clemson University
Clemson, SC
USA
mshuffl@clemson.edu

Christopher Wiese
University of Central Florida
Orlando, FL
chriswwiese@gmail.com

William Kramer
University of Central Florida
Orlando, FL
chriswwiese@gmail.com

Eduardo Salas
University of Central Florida
Orlando, FL
salas@ist.ucf.edu

Abstract: Performance assessment plays an important role in organizational development by informing the development of training programs and identifying employees who are ready for career advancement. Such assessments are largely conducted by supervisors who, formally or informally, evaluate performance. This imposes high resource requirements and there is a loss of standardization and objectivity in these situations. Simulations are better tools for assessing hands-on skills but there is limited guidance for developing them. Furthermore, they are costly to develop. This paper describes both a methodology for developing robust simulations for assessing expertise and skill levels of employees, and a tool for the rapid, low-cost development of such simulations. We focus especially on the results of studies performed to validate the methodology. The results show that this approach holds promise for producing effective criterion measures for task performance.

Traditional measurement of job performance involves a myriad of methods, which are centered on paper and pencil assessments (e.g., declarative knowledge tests, supervisor reports). However, there are instances in which these measures are not the best choice. For example, if the job being assessed is extremely complex and involves a number of tasks that do not necessarily happen on a frequent basis, it would be difficult to determine job performance with a standard test. Therefore, by leveraging the control and structure afforded by simulation environments, the obstacles of capturing complex job performance can be managed. They allow for replication of critical tasks that one might only experience while deployed; they can also provide supervisors with the most comprehensive overview of someone's performance prior to promotion or selection decisions.

There are, however, a number of barriers that arise when developing a simulation-based performance assessment. Simulations are much more expensive to develop than multiple-choice batteries and there are few guidelines for developing them. An assessment simulation must measure relevant skills and the assessment must be valid, and ensuring validity can be a challenge. Care must be taken to ensure that the simulations measure job skills and not the ability to use computers or the ability to game the system. Therefore, developing an evidence-based approach that follows from performance assessment theory and practice is of central importance to a valid simulation-based performance assessment (Mislevy, Steinberg, Breyer, Almond, & Johnson, 1999).

Simulation methodology and development

The simulation development methodology that we followed was discussed in detail in (Ramachandran, Ludwig, Salas, & Rosen, 2008). Here we provide a brief overview. The methodology prescribes the following eight-step procedure for simulation development.

Step 1: Define what needs to be measured.
Step 2: Develop a sampling strategy.
Step 3: Generate scenarios with embedded events and measurement tools.
Step 4: Decide on an appropriate scaling technique and encode in a measurement tool.
Step 6: Administer the simulation and measurement tools to a developmental sample.
Step 7: Evaluate the scenarios and measurement tools.
Step 8: Optimize the scenarios for performance assessment.

The types of simulations we used and the tool for developing the simulations are described in (Ramachandran, Sincoff, Shuffler, Wiese, & Salas, 2011). We developed simulations designed specifically for rapid development, using S-CAP, a tool that provides an abstract simulation model that can be rapidly instantiated for specific domains. The S-CAP authoring tool facilitates the creation of simulation-based assessments that can be run using an Internet web browser as a player interface. No programming skills are required for its use, making it a tool that is usable by subject-matter experts in various content areas. The simulations produced by S-CAP are Adobe Flash applications and are platform independent. Users interact with the simulations using menus and pop-up multiple-choice questions. S-CAP's visuals are largely restricted to static, two-dimensional images to represent scenes. However, it has the capability to play videos and audio when necessary. S-CAP keeps track of a user's actions within a simulation, matching these with an assessment key in order to automate performance assessment.

The authoring tool facilitates content creation by providing a simulation template that can be filled out to create new assessment scenarios. The template defines a collection of abstract simulation elements that represent the logical components of a simulation. S-CAP simulations consist of the following main components: scenes representing the locations in a simulation, a set of simulation items or entities, a set of actions that can be performed on each item, and a set of events that are automatically triggered under specified conditions. The elements are tied together by two other important components: conditions and effects. Conditions are descriptions of specific simulation states, and effects are the specifications of the changes to simulation states resulting from actions or events. Finally, each simulation has an associated set of assessors that specify the rubric for evaluating performance. The process of creating a new scenario involves instantiating each of these components. The authoring tool mirrors the template and includes editors for specifying the details of each component.

Once a simulation ends, the system scores the user's performance and presents a report card as shown in Figure 1. S-CAP uses a goal-based assessment to overcome the challenges of automatic performance assessment in simulations with multiple solution paths. A goal represents a specific set of conditions that must be true in a state. To assess performance, scenario authors specify a set of goals that must be met at different points in the simulation. Each goal typically has a set of cues that state when the goal is active. This enables S-Cap to represent sequencing information, e.g., a medic must perform various assessments on a patient before starting any treatments. Thus, the assessment goal will serve as a cue to treatment goals so that the latter will only become active once the cue goals are satisfied. This goal-based assessment approach is described in detail in (Ramachandran et al., 2011).

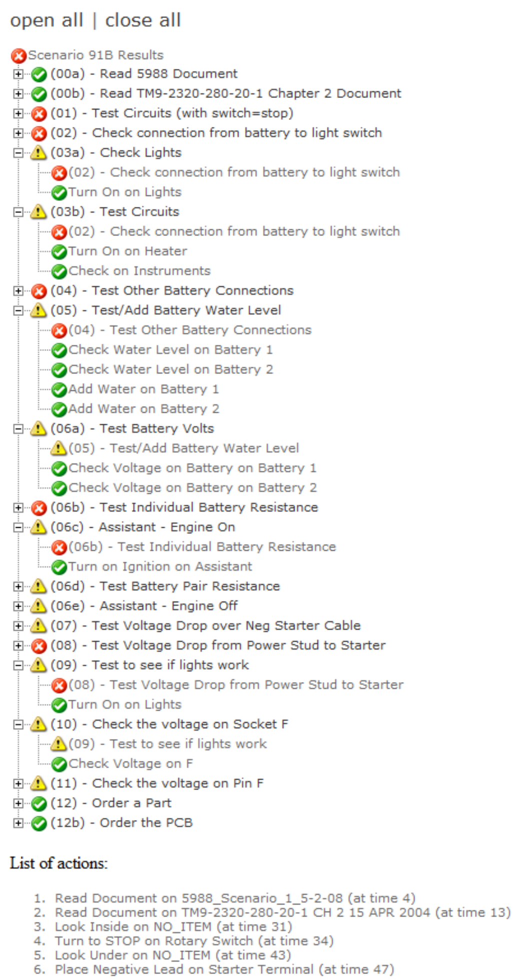

Figure 1: Performance review at the end of a simulation

Going back to the report card generated by the simulation using the goal-based approach (Figure 1), the hierarchically arranged entries show the goals (the second-level Items in the hierarchy) specified in the scenario assessment criteria and the conditions that satisfy them (the leaf-level entries). A checkmark indicates that the corresponding condition was completely satisfied by the user's actions. An exclamation sign indicates that the goal/condition was only partially satisfied. A cross indicates the goal/condition was not satisfied. The assessment for each goal is a function of the assessment for each of its conditions. A goal is fully satisfied only if all the nodes under it are satisfied. However, partial satisfaction of a goal will lead to partial credits, even if the node is marked with a cross in the report card. All extraneous actions performed by the user are not captured in this scorecard. Thus, according to the report card in Figure 1, the user has completed all the conditions needed to achieve the goal "Read 5988 Document" but has not satisfied any of the conditions required to satisfy the goal "Check connection from battery to light switch." Though the user has met the condition "Turn on Lights" for the goal "Check lights," it is still evaluated as partially completed because the user did not successfully complete the goal's precondition, namely "Check connection from battery to light switch."

The simulation currently reports two scores based on an assessment of user actions. The first score measures the extent to which the user followed the expected procedure specified by the author, and is calculated as a percentage of the goals and conditions that were satisfied (including full credit scores and partial credit scores). The second score measures the user's efficiency as the inverse ratio of the number of actions performed by the user and the minimum number of actions required to complete the expected procedure specified by the author. This score does not include an assessment of whether the user met the terminal goals of the simulation or if indeed the scenario was completed to success. Therefore, this score must be used in combination with the first score measure to assess overall performance.

Army case studies

An important goal of this work objective research effort was to study the effectiveness of the simulation development methodology and the S-CAP framework by developing two assessment simulations each for four Army Military Occupational Specialties (MOSs).

We selected the following four Army MOSs as test cases for our approach: 91B, 11B, 68W, and 25U. Diversity of task types, the types of knowledge and skills required for job performance, and the availability of SMEs were the criteria used for MOS selection. It was considered important to select job categories that required a range of cognitive skills; the chosen MOSs cover areas such as equipment troubleshooting and repair, emergency first response, and giving Call For Fire orders. Finally, SMEs are an important source of knowledge to inform the simulation development process and therefore their availability was a significant factor in our consideration.

For each MOS, we followed the eight-step simulation development methodology outlined above. As a first step, we identified documents such as field manuals, training manuals, and operating manuals to serve as sources of content information about the tasks to be assessed. We also identified SMEs to provide details on the steps of identifying critical skills and the appropriate assessment measures and scaling techniques as well as for developing scenarios for assessing the identified critical skills. The SMEs also provided initial validations of the implemented scenarios by viewing demonstrations and providing feedback.

For each MOS, we developed a questionnaire to guide the SME interviews. We received extensive SME input for the 91B, 11B, and 68W MOSs. We interviewed two SMEs for 91B, two for 11B, and one for 68W. SME support was limited for 25U. Although we did receive information from two or more SMEs, our contact with them was indirect and limited to email communications. We had just a few opportunities to conduct an interactive interview. This limited our ability to develop detailed scenarios for the 25U MOS.

The information provided by SMEs was augmented by references to field manuals and equipment operating manuals. The latter contained the details of the procedures and protocols to be followed by users in their jobs and served as the source of information for scenario design. We followed an iterative process of interviewing SMEs and developing scenario versions using technical manuals as references. Three or four such iterations were typically needed in order to establish detailed scenario specifications that could then be implemented.

Developing the media for the simulations was a significant component of the process. We have long recognized that media creation can produce a resource bottleneck in the simulation development process in terms of both cost and time. This is the primary reason S-CAP limits itself largely to simulations that use 2-dimensional graphics and requires only limited amounts of video/audio/graphics. This focus on minimal media largely mitigated the problem of media development bottleneck but still led to some delays. This was especially pronounced for the 91B simulations, where we required photographic images of Army HMMWV vehicles. Obtaining the images required permissions from multiple levels of command within the Army and caused significant delays. As a result, a decision was made to use digitally synthesized images for the remaining MOSs. This helped us avoid organizational delays. However, creating rendered graphic images is slower than photographing actual equipment. Despite this, we recommend the rendered image approach because:

1. Organizational delays are higher sources of risk as they are often unknown and unpredictable, whereas the time required to develop any kind of digital media can be estimated reliably.
2. The images of most equipment and parts are available on the Internet or in technical reference documents and can be used as a reference for generating digital renderings.

Finally, the methodology prescribes that the simulations must first be validated against an assessment sample before being evaluated for validity on a larger sample of population. Since we did not have large numbers of participants available, we were unable to perform the two-stage validation for any of the four MOSs. As such, the validations we performed (discussed below) were essentially pilot validations on an assessment sample. The two-step validation approach prescribed in the methodology is highly recommended, based on our experience.

Validation

Validation is an essential step in the development process to ensure that the resulting simulations are practically useful and accurately measure what they intend to measure, especially in terms of performance assessment. Thus, the final aspect of this research was the undertaking of empirical validation efforts for each of the simulations developed for the four MOSs (91B, 11B, 68W, 25U). These validations were conducted with active duty Soldiers in the relevant MOSs in order to ensure that they most accurately reflected the population that would be utilizing such simulations for performance assessments.

We used concurrent validation that involves the administration of the performance assessment along with the simultaneous collection of some other type of already validated measure of performance or experience (Nunnally & Bernstein, 1994). For example, a concurrent validation could involve having Soldiers of different skill levels (in order to validate across the full range of performers) take both the simulation-based performance assessment tool that we have created as well as a validated measure of performance on the same task (e.g., a pencil/paper measure) or some other measure of performance (e.g., supervisor ratings). These scores can then be compared to determine if there is similarity in performance across both measures (e.g., those who score low on validated paper/pencil measure also score low on new assessment). If there is similarity, then the new assessment tool can be considered valid. For the purposes of our validation, the comparison metrics were past performance experience as rated by the Soldiers. While a more objective measure such as supervisor ratings is more desirable for this type of validation, given data collection constraints (i.e., no available objective measures), the use of prior experience in relation to the tasks being assessed is an acceptable substitute as it is a predictor of future performance (Schmidt & Hunter, 1983).

Our overarching hypothesis for each of the validations was that Soldiers with more experience and longer tenure in their MOSs would achieve higher scores on the simulation assessment than those with less experience. The following sections provide more detail regarding the methodology utilized across each of the four sets of validations as well as the presentation and discussion of each set of results for the validations.

Methodology

Sample Two-hundred-sixty-eight U.S. Army Soldiers were recruited for participation in the four validations via research umbrella weeks held at different locations around the United States. Soldiers were recruited based on their MOSs for each of the specific simulations. Larger samples were planned in order to do a two-stage validation of initial piloting and subsequent actual validation based on feedback during the pilots, but due to the tempo of current military operations and data collection constraints, this was not possible. Additionally, while ideally the validations would be conducted using Soldiers of different skill levels (i.e., Skill Level 1–2) in order to make comparisons among these groups, due to data collection restrictions only Soldiers of Skill Level 1 (E1–E4, which represent a wide range of hands-on task experience) were available. The results section provides a brief description of the demographics for each of the four validations.

Measures For each of the four MOSs, a survey was created that assessed basic demographics (e.g., age, gender, MOS) as well as previous experience that could be used as a point of comparison for the validations (e.g., tenure, skill, and experience in performing tasks being assessed, years since Advanced Individual Training or AIT).

Previous Experience As objective performance metrics were not available for these data collections (e.g., supervisor ratings, previously validated tests), previous experience was utilized as a point of comparison. As previously discussed, this type of data has been shown to successfully predict job performance (Schmidt & Hunter, 1983) and therefore was expected to also be related here to performance on the assessment tools. For this validation effort, we utilized data collected from Soldiers regarding their general military experience as well as their specific levels of experience in areas relevant to the skills being tested. For example, in addition to general tenure in the Army, for the 68W MOS questions were asked regarding experience in triaging patients, performing care under fire, and treating gunshot wounds.

Control Variables Several control variables were also assessed in the demographic surveys. This included controlling for previous computer experience, video game experience, gender, and age.

Performance Score As previously discussed, after completing the assessment scenarios, participants received scores based on the measures described earlier. The exact composition and range of each of these scores varied by MOS as a function of the scenario design as some involved a greater number of steps or procedures that had to be completed to achieve a higher score.

Procedure After obtaining informed consent, participants were first asked to complete a basic demographic questionnaire. Participants were next given a 15-minute block of instruction on the simulation and its modes of interactivity. Additionally, they were given 15 minutes to explore a practice simulation and learn the controls of the interface.

After the initial training, participants were given one of the two assessment scenarios being validated and asked to do their best to complete the task (with the exception of the 91B MOS, where participants only completed one scenario). Once the simulation was over, scores were recorded according to the scoring mechanisms coded within the simulations. This was then repeated for the second scenario. Following the completion of both scenarios, participants were debriefed and asked to provide their reactions to the simulation for face and utility validation purposes. In trials with two simulations, say Simulation A and Simulation B, the ordering of the two simulations was varied so that half the participants worked first on Simulation A first and the other half worked first on Simulation B.

We first describe the details of the data collection for each MOS. The validation results are summarized in Table 1.

91B Data Collection

Data were collected from 139 U.S. Army Soldiers, all of whom had the MOS of 91B. In order to determine if level of experience with conducting HUMVEE electrical repairs was significantly related to simulation assessment scores, a hierarchical multiple regression analysis was conducted.

11B Data Collection

Data were collected from 81 U.S. Army Soldiers at two different Army posts in the United States. For this set of scenarios, experience questions specifically focused on previous experience with performing calls for fire and utilizing different tools for measuring the distance to targets, as these were the primary skills being assessed in this simulation. For this MOS, two scenarios were developed (referred to as Shift and Grid), and therefore both were validated. In order to prevent order effects, the order in which participants completed the scenarios was alternated so that there was a relatively even divide in terms of the number of participants completing the Grid scenario first or the Shift scenario first.

68W Data Collection

Data were collected from 29 U.S. Army Soldiers from 68W. For this MOS, experience questions specifically focused on previous experience with performing care under fire as this was the primary skill being assessed in this simulation. Similar to the 11B simulations, there were two scenarios that were validated, one regarding treating a bleeding wound and the other in regard to triaging patients and treating a gunshot wound. Again, in order to prevent order effects, the order in which participants completed the scenarios was alternated, so that there was a relatively even divide in terms of the number of participants completing the Bleeding Wound scenario first or the Triage/Gunshot Wound scenario first.

Given the sample size for this data validation effort, this was essentially a pilot validation as opposed to a true validation effort. Thus, the data collection efforts provided rich context for improving the scenarios and the scoring mechanisms but offered less in terms of true validation data. Empirical results are reported for each but should be interpreted with extreme caution.

25U Data Collection

Data were collected from 19 U.S. Army Soldiers, 11 of whom had the MOS of 25U while the remaining eight were elsewhere within the signal career field. This variance in MOSs was due to difficulty during the data collection in locating Soldiers in the MOS specified. However, this provided variance in terms of the skill and experience levels of the participants, so these MOSs were all included in the analyses. Data collections were conducted at one Army post in the United States. For this MOS, experience questions specifically focused on previous experience working with the Defense Advanced GPS Receiver (DAGR) and Single Channel Ground and Airborne Radio System (SINCGAR) communication systems and repairing communication systems in general as these were the primary skills being assessed in this simulation.

Similar to the previous two sets of simulations, there were two scenarios that were validated, one regarding repairing a DAGR system and the other in regard to repairing a SINCGAR system. Again, in order to prevent order effects, the order in which participants completed the scenarios was alternated so that there was a relatively even divide in terms of the number of participants completing the DAGR scenario first or the SINCGAR scenario first. In designing these scenarios, there was very little access to SMEs. Therefore, this data validation was also primarily a pilot validation, especially as there was little time to review the developed scenarios prior to the commencement of the validation efforts.

To summarize, the overall results for each MOS are summarized in the following table.

MOS	Sample Size	Result Summary
91B	139	1. Prior experience did predict performance assessment scores, above and beyond the control variables (ΔR^2=.03, F(1,133)=4.426, p<.05). 2. Specifically, electrical experience was positively and significantly related to assessment scores (β = .18, p<.05). Furthermore, none of the control variables (computer experience, video game experience, age, gender) predicted

		assessment scores. 3. None of the control variables (computer experience, video game experience, age, gender) predicted assessment scores.
11B	81	1. Regression results for one scenario illustrated that prior experience predicted performance assessment scores, above and beyond the control variables (ΔR^2=.11, F(1,76)=4.426, p<.05). For the second scenario, there were no significant predictors of performance. 2. Computer experience and video game experience did not predict assessment scores; however, age was a significant predictor when included with grid experience (β = -.35, p<.05). 3. For the second scenario, time since AIT did appear to be significantly related to performance (Table 5); ΔR^2=.08, F(1,76)=7.10, p<.05; β = -.32, p<.05).
68W	29	1. For one of the scenarios, past experience was negatively and significantly correlated with assessment scores (β = -.44, p<.05). No significant relationship between prior experience and assessment scores was found (ΔR^2=.08, F(3,20)=.65, *p*=.59) for the second scenario. 2. Age and video game experience did not predict assessment scores; however, computer experience was a significant predictor when included with gunshot experience (β = .60, p<.05). 3. For the second scenario, there was a negative correlation between pay grade and assessment scores (β = -.50, p<.05). This relationship with pay grade may mean that those with higher pay grades are less likely to be performing this task and therefore may not have performed as well, while those with lower pay grades (for whom the scenario was designed) have performed this task more recently and may therefore score better. Additionally, it may be the case that those with lower pay grades are more likely to have computer/video game experience due to their age and therefore are more adept at adapting to the computer simulation demands.
25U	19	1. For both of the scenarios validated, regression results illustrated that prior experience did predict performance assessment scores above and beyond the control variables. 2. None of the control variables (computer experience, video game experience, age, gender) predicted assessment scores.

Table 1: Summary of results

Discussion of Results

The objective of this effort was to study the effectiveness of job performance assessment using simulations created by following the methodology described in the section titled Simulation methodology and development and using the S-CAP framework. What can we conclude from the validation study results? Of the four MOSs studied, two resulted in a positive correlation between prior experience and the simulations' performance assessment scores, thus providing confidence that the simulations indeed predict competence. For one of the MOSs, a positive correlation was also observed but for only one of the two assessment simulations. By contrast, for one of the MOSs, the assessments scores from the simulations were negatively correlated with the level of prior experience. The simulations for this MOS were also the most open-ended, giving the user a high degree of freedom for exploration. Thus, it is possible that the effectiveness of the simulations at performance assessment was confounded by difficulties in learning how to operate them. An initial pilot study to uncover and correct usability issues could have been helpful. We were unable to conduct such studies due to logistical challenges.

The methodology does prescribe an initial pilot sample prior to the full validation. During many of the data collection trips, useful input was gained regarding the face validity of the systems, potential problems and drawbacks related to the technical aspects, and feedback regarding how steps might be approached differently by different SMEs. Furthermore, it was also interesting to note that some Soldiers had their own methods of accomplishing tasks, which, while they led to the same answer at the end of the performance period, did not follow the same steps as laid out in the technical manuals and similar documents. Therefore, being able to test with a pilot sample of at least 20–25 Soldiers would have been ideal in each situation. This would have helped adjust the overall scenarios prior to a full-scale validation. Even being able to test with a smaller focus group would have given additional guidance, particularly in terms of the reproduction of skilled hand movements to a computer display format. Such guidance could have helped the 68W simulations as these, of the eight we developed, offered the most tools and choices of actions to the users.

Finally, ideally this validation would have been conducted using less subjective measures of previous experience. In the future, obtaining supervisor ratings or some other source of performance would provide a better, more objective point of comparison. Given the limits on time and accessibility, asking only a short set of items regarding previous experience may have limited our ability to truly capture the extent to which participants had experiences relevant to the tasks being performed.

Overall, even with these limitations, it does appear that the methodology proposed and the related tools for simulation development do have promise for the authoring of performance assessments. This is particularly important given the current lack of objective, reliable assessments for many existing MOSs and related tasks. Indeed, we consistently heard throughout the validation effort that these types of tools were something that could be very useful for many MOSs, given the lack of anything similar currently available. It is hoped that the lessons learned during this research will provide future endeavors with a clearer direction for developing such simulation tools for performance assessments.

In terms of the process, the S-CAP tool did facilitate rapid simulation development. The most time-consuming aspects of the development process were defining the scenarios in detail and developing media. Authoring the simulation and the assessment rules only required one or two weeks per scenario. This points to the suitability of S-CAP as a cost-effective tool for developing simulation-based job performance assessment batteries.

Lessons learned

The experience of applying this process on a range of real domains has led to the following lessons:

Lesson 1: Both SMEs and published documents are important sources of information for scenario development. While technical manuals provide details about how specific tasks are to be performed, they do not provide information on which tasks and skills are critical to an MOS or what skills differentiate between experienced and inexperienced workers. This knowledge is critical to the development of assessment simulations. Therefore, interviewing SMEs is of vital importance. However, SMEs are time constrained themselves and do not have the time to provide detailed scenario information. Thus, the best uses of SMEs are to help identify critical skills and tasks and to help develop the scenario focus and outline. Technical manuals can then be used to fill in the scenario details. Multiple iterations of SME interviews and technical manual reference will be required to get a scenario specified in sufficient detail for implementation.

Lesson 2: Offline communications with SMEs are not as effective or useful as interactive conversations and interviews. We had greater success in cases where we were able to interview SMEs interactively over the phone. Personal, face-to-face interactions are not necessary. Telephone conversations are effective. However, offline communications over email alone are not sufficient and can significantly delay the development process.

Lesson 3: Creating digitally rendered images based on real equipment and people is preferable to trying to acquire photographic images. It is often difficult to gain access to military equipment and personnel as it can require permissions from multiple levels of command. However, images of most devices can be found openly and can serve as references for generating digital image renderings. Such an approach may result in fewer delays.

Lesson 4: Validation on an assessment sample is a vital step before large-scale evaluations of the validity of assessment simulation can take place. There is no substitute for having target users run through the simulations to discover aspects of them that need to be changed or improved. User interface, scoring, and the accuracy of steps all need to be evaluated using an assessment sample. The purpose of this validation is not to assess the accuracy and reliability of the simulations in predicting expertise but in uncovering problems with the simulations that may interfere with their effectiveness. Therefore, the focus of the validation with the assessment sample must be in discovering such problems early in the process. If resources permit, multiple evaluations on assessment samples are recommended.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily represent the official policy or position of the Department of Defense or the US Army Research Institute, which funded this work under contract number W91WAW-09-C-0134.

References

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing simulation-based performance assessment. *Computers in Human Behavior*, *15*(3), 335-374.

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. F. (1967). *Psychometric theory* (2). McGraw-Hill New York.

Ramachandran, S., Ludwig, J., Salas, E., & Rosen, M. (2008). A methodology for simulation-based job performance assessment. *Proceedings from The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*.

Ramachandran, S., Sincoff, E., Shuffler, M., Wiese, C., & Salas, E. (2011). An Examination of the Effectiveness of Simulations as Performance Assessment Instruments. *Proceedings from The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*.

Schmidt, F. L., & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, *68*(3).