An Examination of the Effectiveness of Simulations as Performance Assessment Instruments

Sowmya Ramachandran[1], Erik Sincoff[1]

[1]Stottler Henke Associates

San Mateo, CA

sowmya@stottlerhenke.com, esincoff@email.com

Marissa Shuffler[2], Christopher Wiese[2], & Eduardo Salas[2]

[2]University of Central Florida

Orlando, FL

mshuffler@ist.ucf.edu, chriswwiese@gmail.com, esalas@ist.ucf.edu

ABSTRACT

Job performance measurement is of critical importance to any organization's health. It is important not only to recognize and reward good performance, but also to groom future leaders. Developing effective assessment techniques that are valid, effective, and fair is an ongoing challenge. Assessing factual knowledge using multiple-choice test batteries is relatively inexpensive and is therefore widely used for performance evaluations. Hands-on assessment is the most effective way to assess task proficiency but is very resource intensive and expensive. Computer-based simulations provide an alternative where users can be assessed in the context of skill application under controlled conditions. However, simulations are expensive to produce and maintain. Validated guidelines and methodologies are needed to help organizations develop effective assessment simulations. Our research has the objective of developing a prescriptive methodology and a framework for rapidly creating and deploying simulations. Following up on our earlier report on the methodology itself, in this paper we present the results of applying this approach to assess the performance of light-wheeled vehicle mechanics. Further, we discuss the validity of the simulation developed using this methodology with respect to predicting job performance. We also discuss the benefits and limitations of this methodology and the class of tasks for which it may be best suited.

ABOUT THE AUTHORS

Dr. Sowmya Ramachandran, PhD**,** is a research scientist at Stottler Henke Associates, a small business dedicated to providing innovative Artificial Intelligence solutions to real-world problems. Dr. Ramachandran's interests focus on intelligent training and education technology including intelligent tutoring and intelligent synthetic agents for simulations. She is also interested in issues of motivation and metacognition. Experience with military and private industry gives Dr. Ramachandran a unique perspective on the needs and requirements of the ultimate end-users and their constraints. She contributes expertise in AI, instructional systems, probabilistic reasoning, and knowledge management. She has developed Intelligent Tutoring Systems for a range of topics including reading comprehension, high-school algebra, helicopter piloting, and healthcare domains. She has participated in workshops organized by the Learning Federation, a division of the Federation of American Scientists, to lay out a roadmap for critical future research and funding in the area of ITSs and virtual patient simulations. She has developed a general-purpose authoring framework for rapid development of ITSs, which is currently being used to develop an intelligent tutor to train Navy Tactical Action Officers. She has also developed tools and technologies for training emergency first responders.

Mr. Erik Sincoff received his Masters degree in Computer Science with a focus in Artificial Intelligence (natural language processing, intelligent agents, and collaborative filtering,) and computer-based education (human-computer interaction, computer graphics and tutoring systems.) While at Teknowledge Corporation from 1994 to 2005, he gained extensive experience with Education and Training systems. This experience entailed the DoD sponsored Computer Aided Education and Training Initiative and the Courseware Conversion Factory project via the NIST ATP program. While at Stottler Henke, Mr. Sincoff has worked on the design and implementation of various intelligent tutoring systems in the domains of counterterrorism, emergency medical response, and military tactical controls.

Marissa L. Shuffler, MA, is a doctoral candidate in the Industrial and Organizational Psychology program at The University of Central Florida (UCF). She earned a B.A. from Pfeiffer University with a triple major in Psychology, Organizational Communications, and English, and a M.A. in Industrial/Organizational Psychology from George Mason University. Ms. Shuffler is a graduate research associate at the Institute for Simulation and Training where her research interests include individual and team training, leadership, intercultural collaboration, multi-team systems, and decision-making/adaptation, with an emphasis on military environments. She is student scientific lead on several major projects investigating leadership training and development for leaders operating in complex, multicultural, distributed environments. Additionally, as an external consultant Ms. Shuffler has conducted research for the U.S. Army Research Institute, the Department of Labor, the Center for Army Leadership, the Department of Homeland Security, and the U.S. Air Force. Her work to date includes over 70 publications and presentations/posters at professional conferences.

Christopher W. Wiese, B.S. is a doctoral student in the Industrial and Organizational Psychology program at the University of Central Florida (UCF). He earned his B.S. in psychology with a minor in Business Management from UCF in 2007. Mr. Wiese is a graduate research assistant for both the Team and Workforce Development lab and the Institute for Simulation and Training where his research activities include the topics of individual assessment, cognition in teams, technology enabled communication, leadership, multicultural collaboration, and skill transparency in assessment centers. He has conducted research in association with U.S. Army Research Institute and is involved in two Multi-University Research Initiatives (MURI).

Eduardo Salas is Trustee Chair and Professor of Psychology at the University of Central Florida where he also holds an appointment as Program Director for the Human Systems Integration Research Department at the Institute for Simulation and Training. Previously, he was the Director of UCF's Applied Experimental & Human Factors Ph.D. Program. Before joining IST, he was a senior research psychologist and Head of the Training Technology Development Branch of NAWC-TSD for 15 years. During this period, Dr. Salas served as a principal investigator for numerous R&D programs, including TADMUS, that focused on teamwork, team training, decision-making under stress and performance assessment. Dr. Salas has co-authored over 300 journal articles & book chapters and has co-edited 19 books. His expertise includes assisting organizations in how to foster teamwork, design and implement team training strategies, facilitate training effectiveness, manage decision making under stress, and develop performance measurement tools. Dr. Salas is a Fellow of the American Psychological Association, the Human Factors and Ergonomics Society, and a recipient of the Meritorious Civil Service Award from the Department of the Navy.

An Examination of the Effectiveness of Simulations as Performance Assessment Instruments

Sowmya Ramachandran[1], Erik Sincoff[1]

[1]Stottler Henke Associates

San Mateo, CA

sowmya@stottlerhenke.com, esincoff@email.com

Marissa Shuffler[2], Christopher Wiese[2], & Eduardo Salas[2]

[2]University of Central Florida

Orlando, FL

mshuffler@ist.ucf.edu, chriswwiese@gmail.com, esalas@ist.ucf.edu

BACKGROUND

The challenge of hiring, retaining, and training effective workers has long been an overarching concern for organizations. This is markedly so for organizations such as the military where people's lives depend on job competence, both their own and that of others who work with them. Getting an accurate picture of an employee's competence is therefore essential for the success of such organizations.

Among the number of prevailing approaches to performance assessment, the most common is the use of test batteries with multiple-choice questions that have been carefully designed and validated (Campbell, Keenan, Moriarty, Knapp, and Heffner 2004). These types of instruments have strong theoretical models that guide their development and validation so that the tests can be used with a high degree of confidence that they accurately measure the skills/expertise of interest. Validation techniques have also been developed to demonstrate that the relative performance of individuals on the tests accurately reflects the differences in their expertise levels.

Multiple choice tests are a fine choice for assessing factual knowledge, but they are not necessarily appropriate for measuring the ability to apply skills on the job. The problem of inert knowledge is well-known and well-documented (Schank 1995). Inert knowledge reflects the phenomena where people possess sufficient factual knowledge but lack the proficiency to apply this knowledge to solve real problems. For example, a light-wheeled vehicle mechanic may have knowledge of all the parts of a vehicle and how they connect with each other, but may lack the practical skills for troubleshooting a defective vehicle efficiently.

On–the-job assessments can test skills in operational contexts. However, they are resource intensive and typically require a one-to-one ratio between assessors and assessees. Furthermore, it can be difficult to ensure fairness and objectivity in assessment in such settings. This deficiency is sometimes addressed by using two assessors to ensure objectivity but this leads to further increase in resource requirements.

Simulations provide many of the benefits of on-the-job testing by assessing skills in the context of realistic work situations. Simulations typically include automated performance assessment. This overcomes the issues of uniformity and objectivity and eliminates the need for one-on-one time with an assessor. However, there is limited theoretical guidance for their development and validation. Such guidance is critical to ensure that there is a high degree of confidence and trust among stakeholders in the assessments provided by the simulations. 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. In addition to the lack of theoretical guidance, simulations are also considerably more expensive to develop than multiple-choice tests.

Recent research is starting to provide evidence that an infrastructure of prescriptive methodologies and low-cost, rapid authoring simulation tools can be developed that will result in effective performance assessment tools. Following up on our earlier report on one such methodology (Ramachandran et.al. 2008), in this paper we will present case-study of applying this methodology to develop a simulation-based assessment instrument for evaluating the performance of light-wheeled vehicle mechanics. Of primary interest is the validity of the simulation in predicting expertise. This will be discussed in the Section titled **Error! Reference source not found.**. The details of the simulation being validated will be discussed in the Section titled **Error! Reference source not found.**. The paper will conclude with a discussion of the benefits and limitations of our simulation framework and the class of tasks for which it may be best suited.

METHODOLOGY FOR DEVELOPING SIMULATIONS FOR PERFORMANCE ASSESSMENT

Due to the far-reaching impact of job performance assessments, it is important that assessment instruments target critical job skills. It is also important to demonstrate that the simulations do indeed assess the target skills and that the test scores correlate in significant and meaningful ways with actual job skills. Ramachandran et.al. 2008 described a prescriptive step-by-step procedure for developing such simulations. This procedure was developed based on 1) a review of current literature on the design of simulation scenarios and measurement tools as well as the development of selection systems and test items, and 2) practical experience implementing the methodology in developing the prototype simulation. This section just provides a brief overview. A more detailed description is found in Ramachandran et.al. 2008.

The methodology we have defined consists of the following steps:

Step 1: Define clearly what needs to be measured.

Any effective measurement system begins with a clear definition of what is to be measured. This step is accomplished by reviewing pre-existing documents like job-analysis descriptions, training materials, technical manuals, and standard operating procedures to identify key performance measures. Structured interviews with subject matter experts (SMEs) are also necessary. The knowledge obtained from the above sources must be synthesized into a list of competencies and associated performance contexts.

Step 2: Develop a sampling strategy.

It is not feasible to develop scenarios that test every competency or skill required by a job category. It is necessary to find a critical representative sample of skills to address via the simulation. Sometimes several simulations will have to be developed to cover the critical skills. With job categories that are very broad, there will be a trade-off to be made between simulation development resources and coverage of competencies. It is therefore important to have a systematic way of identifying and linking critical skills to simulation scenarios. Second, systematically linking scenario development to the targeted competencies affords the ability to track what competencies have and have not been sampled by the simulation scenario.

Methods of sampling strategies for competencies can focus on time, criticality, and level (Sackett & Laczo, 2003), which can be determined from pre-existing documents or by structured interviews with SMEs.

Steps 3: Generate scenarios with embedded events and measurement tools.

The process of developing simulation scenarios is central to using simulations for selection purposes. Cognitive and behavioral task analysis techniques (e.g., critical decision method, hierarchical task analysis) can be leveraged to sample the range of tasks required and situations encountered for a specific job. For procedural skills, the fundamental outlines of simulation scenarios can often be generated from existing technical and training references. Once an outline of the simulation has been created, the general process involves progressively contextualizing the abstract competencies, using SME guidance to focus on key competencies, using supporting documentation to generate the overall structure of a scenario, and using SME interviews to provide details about each component of the procedural task. Scenario events should be realistic, aim at the appropriate level of difficulty, provide multiple opportunities to display targeted competencies, and sequential dependencies should be avoided in the measurement associated with events (Fowlkes & Burke, 2005).

Step 4: Decide on an appropriate scaling technique and encode in a measurement tool.

Once a scenario with assessment events has been developed, there remains the step of determining how user performance will be scored. Common metrics include either 1) latency from the time some information is provided to the performance of an action that is expected, 2) a dichotomous scoring of whether an action was or was not taken, or 3) a count of 'missteps' before performing the targeted response. Dichotomous scoring is the most straightforward and easy method to interpret in most cases; however, the number of missteps and latency measures are likely more diagnostic in differentiating between different skill levels. Thus, dichotomous scoring is likely to give the simplest measure of basic competence while the other approaches are more likely to distinguish between competence levels at finer levels of detail.

Step 5: Have scenarios reviewed by subject matter experts (SMEs).

It is important to have SMEs review the simulations before proceeding with the other steps in the methodology. Key considerations for evaluation are: 1) whether the simulation is a good representation of the target task(s) being assessed, and 2) whether the simulation tests the critical skills identified earlier as a part of Steps 1 and 2. It is also important to address usability issues during this review.

Step 6: Administer the simulation and measurement tools to a developmental sample.

The simulation should be run with a sample from the intended population of use for validation purposes. Additionally, measurement of this sample's subsequent performance on the job should be conducted. The data thereby collected will allow for validation and optimization of the simulation test.

Step 7: Evaluate the scenarios and measurement tools.

Using the data from the developmental sample, the characteristics of the simulations scenarios and measurement tools can be evaluated. The primary means by which this is accomplished is through correlating simulation scores with other measures of competency. If multiple scenarios are developed to assess a set of competencies, the validation must also establish that they are equivalent by correlating the scores on these tests.

Step 8: Optimize the selection test.

The simulation-based test can be optimized using information from the evaluation of the data gained from the developmental sample. This information can be used to maximize the predictive power of the test (e.g., increase reliability of measurement at the chosen criterion cutoff; increase diagnosticity over ranges of proficiency as needed). As in traditional scale development, test length and predictive power of the test are often at odds with the practical considerations demanding the shortest tests possible. This is the case with simulations as well; using item response theory and psychometric principles of test design, the shortest tests (simulations) can be designed with the highest level of prediction and therefore the most utility in selection.

This general methodology for developing performance assessment simulations can be applied across a range of technologies and situations. It is designed to not simply be tied to assessing one type of skills, but instead is designed as a procedure that can be utilized in a range of contexts. While we have used it in a cost-effective simulation, the general steps for creating the assessment content and measurement tools can be used for both high and low fidelity simulations. Furthermore, this approach to developing performance assessment simulations can be used to assess a range of skills, particularly those that are cognitive and those that involve critical thinking. While these types of simulated performance assessments are not necessarily best suited for psychomotor or perceptual skill testing in low cost situations where fidelity may not be as high and therefore less likely to emulate real life situations, they are very useful in assessing ability and skill in cognitive responses and behavioral reactions to a given situation. Given that these types of cognitive and behavioral response skills can be challenging to assess in declarative knowledge tests, the use of this methodology to create performance assessment simulations can prove to be advantageous across a range of jobs and contexts.

DESCRIPTION OF THE SIMULATION

We next present an example of how this methodology can be used, specifically in military performance assessments. We have developed simulations for two Military Occupational Specialties (MOSs): 91B (Light-Wheeled vehicle mechanic), and 11B (Infantryman). Both use the same simulation framework and therefore have the same underlying representational model. Empirical validation of both simulations is currently underway, with initial results available for the 91B simulation. The next section will provide an overview of the 91B simulation as well as present these initial validation results as evidence of the usefulness of the overall approach.

The 91B Simulation

The 91B simulation is designed to assess the following skills:

- Troubleshoot vehicle and equipment problems
  - Inspect and test equipment and determine the causes of malfunctions
- Use technical references
  - Use resources and references in performing maintenance procedures

The simulation requires the user to troubleshoot and fix a broken High Mobility Multipurpose Wheeled Vehicle (HMMWV or more commonly, the HUMVEE). At the start of the simulation, the user is given a standard form that describes the observed symptom. In this case, the form mentions that the vehicle in question has no electrical activity. With this information, the user's task is to perform diagnostic tests on the vehicle to determine the fault and order a replacement part if necessary. The simulation is targeted at entry-level soldiers at Skill Level 1 (E1-E4s). According to the domain experts interviewed, the extent to which a user follows the procedure is an important assessment criterion.

Table 1 shows the first few steps of the troubleshooting procedure for electrical problems. Space limitations prohibit us from showing the entire procedure in the paper. Nonetheless, these steps should be sufficient for the following description of the simulation. The column labeled "Cue" in the table describes the event that will require a response from the user. The "Targeted Response" column describes the expected response (as specified in the technical manual). The column "Evaluation Metrics" indicates how the user's responses for this step will be evaluated.

Table 1: First three steps in the HUMVEE troubleshooting procedure

Step	Cue	Targeted Response	Evaluation Metrics
1	Mechanic is provided with 5988-E form detailing problems and history of vehicle.	Mechanic selects appropriate technical reference (electrical system for HMMWV M998) from the sources available. -There are different types of HMMWV's. The procedures outlined in this scenario are for the M998 (the basic model). -The major distracting information in this step involves 1) sections of the manuals for other types of HMMWV or other trucks (e.g, if the mechanic selects information on M1044A1, they have not been able to extract the appropriate information from the 5988-E form).	-dichotomously scored (mechanic did or did not access correct reference)
2	Mechanic accesses correct troubleshooting procedure within the technical manual	1. Mechanic turns the rotary switch to STOP. 2. Mechanic tests all circuits to check if they are powered. This step includes a sequence of separate actions to test various circuits controlling lights circuit, starter circuit, and the heater using a multimeter set to measure voltage.	-dichotomously scored (mechanic did or did not test all the circuits)
3	Mechanic is informed that none of the circuits work when the Rotary Switch is in the STOP position.	Mechanic turns the lights on by performing the "Turn Lights On" action on the dashboard of the vehicle.	-dichotomous scoring

Figure 1 shows what the user sees upon entering the simulation. The center area shows the graphic image that represents the world. The simulation provides multiple views to which a user can navigate (Figure 2). The image area in each view has a set of hotspots with associated menus that represent the actions that can be performed on simulation objects (Figure 3).

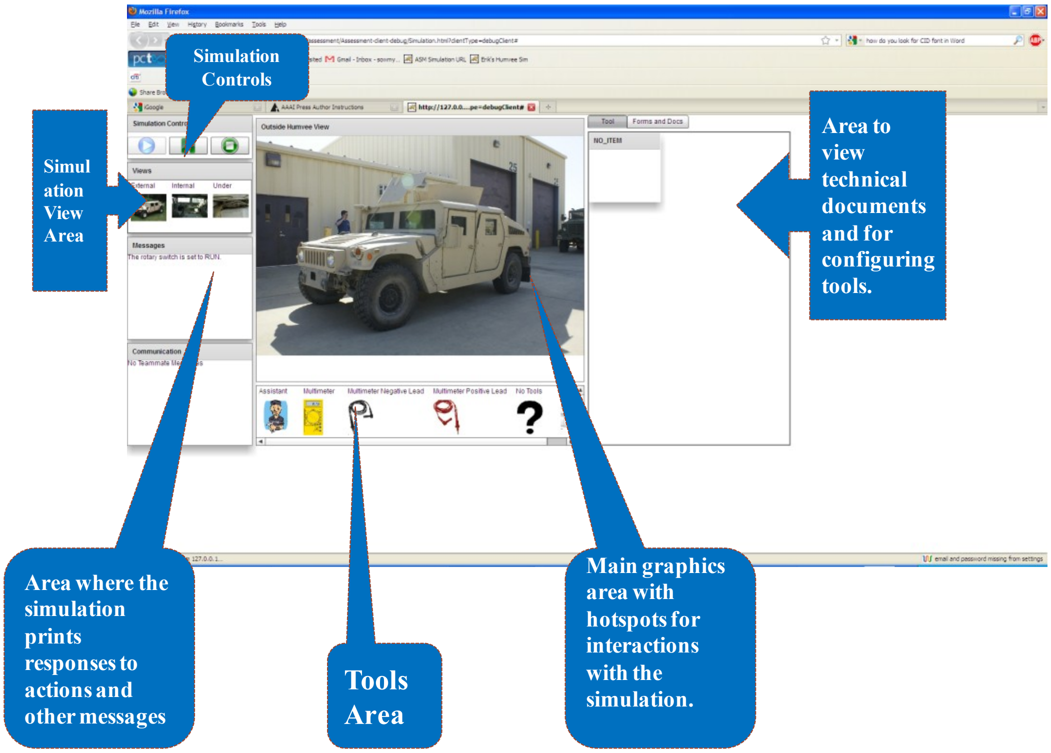

Figure 1: The simulation interface

Figure 2: Simulations can include multiple views

The simulation typically responds to user action with textual messages. We are currently augmenting this to additionally include audio responses.

The simulation also includes a panel of tools (Figure 4) which the user can employ to interact with the simulation and perform tests for troubleshooting. For example, the multimeter tool can be used to find the voltage drop across the terminals of the vehicle batteries.

Finally, the simulation provides a user with hyperlinks to technical manuals for reference. The assessment criteria only require that a user open the hyperlink to the appropriate technical manual. There are no checks to see how the user navigates the document once opened. However, the user's actions are assessed to reflect how closely they follow the troubleshooting procedure specified in the manual.

The simulation ends when the users click on the "End Simulation" button or when they have ordered the correct replacement part, thus meeting the simulation objective.

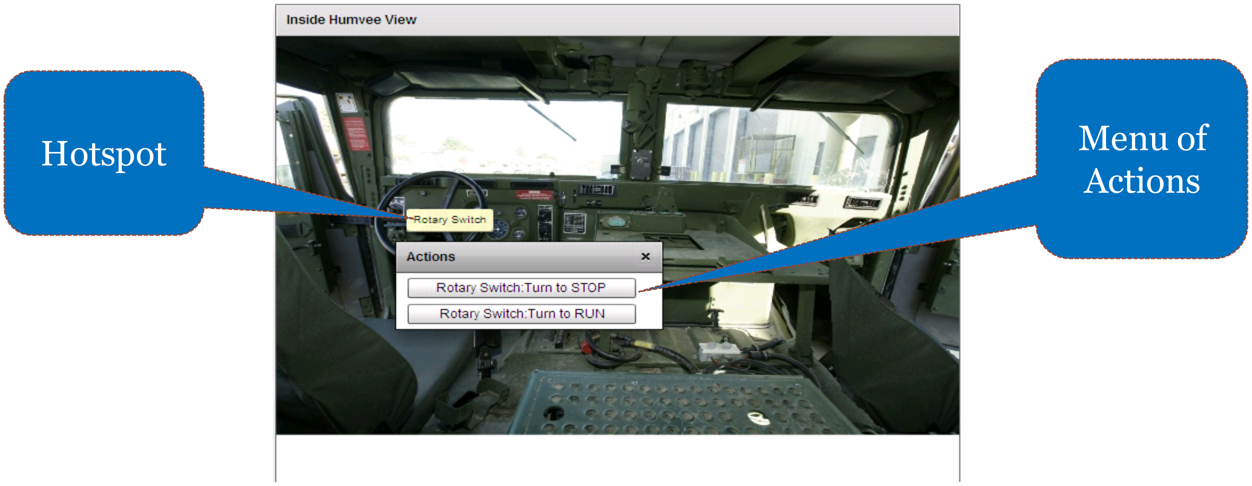

Figure 3: Users interact with the scenario via menu-based actions

Goal-Based Assessment

Once the simulation ends, the system scores the user's performance and presents a report card as shown in Figure 5.

Performance assessment in the context of a free-play simulation presents several technical challenges. First, there may be multiple ways of achieving the same end objective. For example, the voltage drop across two points in an electrical circuit can be measured using a variety of instruments. Specification of such alternative solution pathways can get unmanagebly complex with scoring mechanisms that are based directly on the actions performed. A second challenge is the capture of dependencies and sequencing constraints. Often in procedural tasks, some actions must be performed in sequence whereas others may be performed in any order.

To address these challenges, we have employed a goal-based rather than an action-based representation for simulation scoring. Here, the system has knowledge of the goals that must be met by the user and the simulation conditions that must be true for the goal to be satisfied. For example, rather than state that the user must measure voltage using a voltmeter or an instrument called STE/ICE-R, the assessment conditions specify that the goal of measuring voltage must be met at a specific point in the simulation. Additionally, the specification will state that this goal is satisfied when the simulation variable "Voltage_Measured" is set to true. This variable will be set when the user performs any action that measures voltage (i.e. using the voltmeter or using STE/ICE-R). The benefit of this approach is that the simulation can be extended to include other ways of measuring voltage without changing the assessment criteria. It is sufficient to ensure that the new actions set the simulation condition "Voltage_Measured" to true.

Figure 4: Panel showing the tools available to the user

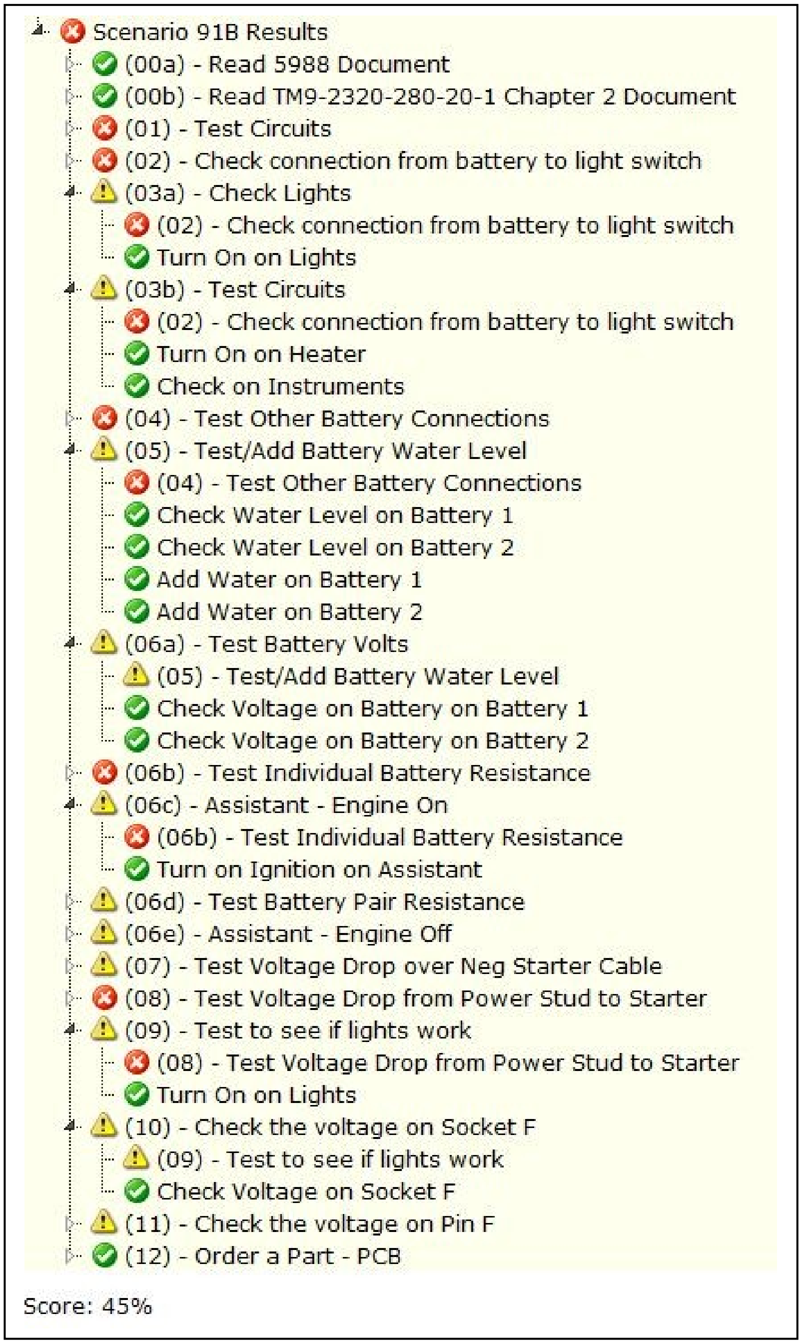

Figure 5: Assessment report at the end of the scenario

Sequencing information is represented in the form of cues. Goals can have other goals that serve as cues. When a goal has a set of cues specified, it will only be evaluated once all the cues have been satisfied. This lets the system enforce an ordering condition on the action. For example, in Figure 5, the goal "Test/Add Battery Water Level" is a cue for the next goal "Test Battery Volts". This means he/she needs to have met the first goal prior to the second in order to get full credit. Thus a user who checks the battery volts without first checking the water level in the batteries will only get partial credit.

Going back to the report card generated by the simulation using the goal-based approach (Figure 5), the hierarchically arranged items 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 items). 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 considered to have been 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. However, this information is used in calculating the efficiency with which the user performed a task. The user's score on the simulation is calculated as a percentage of the goals and conditions that were satisfied (including full credit scores and partial credit scores).

The simulation model consists of a collection of simulation objects that represent the real world objects (e.g., the HUMVEE). Each object has a set of properties that models its state at a desired level of detail. In this simulation, the HUMVEE object has properties that indicate the state of different components (e.g., light working or not working, the fluid level in the batteries). Actions are events that query an object's property list and return the associated values. They also specify the message that should be sent back to the user in response. Actions can additionally change the value of a property and thus effect a change in the simulation state. Events can also be initiated independently by the simulation without any input from the user. For example, simulations can be authored to include timers that cause a state change at certain times. Much of scenario authoring consists of defining objects and their properties, and defining actions and events. Thus, we see that the simulation framework is fairly simple, yet it facilitates the creation of dynamic simulations that go beyond just branching based on user input.

VALIDATION RESULTS

Validation of simulations is an essential step in their development to ensure they are practically useful. Thus, an empirical validation effort was undertaken with the 91B simulation assessment tool developed using the prescriptive methodology and rapid simulation-development tools. A form of concurrent validation was utilized to validate the 91B scenario previously described. Concurrent validation involves the administration of the performance assessment along with the simultaneous collection of some other type of measure of performance or experience that is already validated (Nunnally & Bernstein, 1994). For this validation effort, we utilized data collected from Soldiers regarding their tenure in their MOS as well as their specific level of experience in HUMVEE electrical repairs. It was hypothesized that Soldiers with more experience in conducting HUMVEE electrical repairs would receive better scores on the simulation assessment than those with less experience. While ideally this concurrent validation would be conducted using Soldiers of different skill levels (i.e., Skill Level 1-4), due to data collection restrictions, only Soldiers of Skill Level 1 were available.

Sample & Procedure

In order to test the above hypothesis, data were collected from 139 U.S. Army Soldiers, all of which had the MOS of 91B. Data collections were conducted at three different Army posts around the United States. Each of the participants engaged in a 2-hour assessment session with a computer-based simulation. All participants completed a brief demographic survey which collected information regarding their general military experience, experience in electrical repairs of HUMVEES, and computer and video game experience.

Participants were next presented with a 15-minute presentation on the simulation and its modes of interactivity. Additionally, they were given 15 minutes to explore the tutorial simulation and learn the controls. After the initial training they were given the 91B assessment scenario being validated and asked to do their best to complete the task. Once the simulation was over, scores were recorded according to the scoring mechanisms described previously, and participants were debriefed and asked to provide their reactions to the simulation for face validation.

Results

In order to determine if level of experience with conducting HUMVEE electrical repairs was significantly related to simulation assessment scores, a correlation analysis was conducted. Correlation results showed that electrical experience was positively and significantly related to assessment scores ($r = .17$, $p<.05$). Furthermore, correlations between computer experience and video game experience and assessment scores were non-significant ($r = .085$, $p = .16$, $r = .081$, $p = .17$, respectively). Thus, the scores were not a result of simply being more experienced in computer usage.

Overall, these findings provide support for the validation of this assessment tool as a useful method for assessing 91B HUMVEE repair performance. Although additional validation is needed for the scenarios developed using this methodology for the other MOSs, these findings provide an initial first step towards validating our approach.

CONCLUSIONS

The simulation framework is designed for cost-effective production of performance assessment simulations. The visual fidelity and even the underlying world model fidelity have been restricted in order to control the cost of production. Nonetheless, these simulations call upon expertise that goes beyond recall of factual knowledge. The hypothesis is that these types of simulations are sufficient for a large class of tasks performed in the military and will provide a more robust assessment than current methods employed.

The results discussed above showed that the simulation's assessment correlates positively with a user's experience with the task, demonstrating that the simulation approach is effective for this task. We will have further opportunities to study this hypothesis as we develop and validate simulations for three additional MOSs as a part of this effort.

While this approach places limitations on scenario design, we hypothesize that the simulation model will be effective for job performance assessments for a number of domains that meet the following conditions:

1. The goal is to assess cognitive skills and not perceptual or psychomotor skills.

2. The underlying model can be effectively represented by objects and discrete variables that are controlled by discrete events. It is not required that the underlying model be simple, but that the discrete event and variable based representation be adequate for reliably measuring the critical skills. The HUMVEE is a complex piece of machinery that involves sophisticated principles of energy conversion. However, the event-based representation of our simulation was sufficient for testing the critical KSAs.

Our results for the 91B simulation are a step towards confirming this hypothesis. We plan to develop and validate simulations for three other MOSs for further confirmation.

While the validation study compares the simulation-based assessment approach with independent measures of expertise like experience level and background, we have been unable to directly draw comparisons to other assessment approaches such as multiple-choice tests. This is because the MOSs under consideration do not use such tests to evaluate job performance. Going forward, comparisons with other assessment approaches will be important to establish the relative costs and benefits of the simulation-based approach to job performance assessment.

ACKNOWLEDGMENTS

This material is based upon work supported by the Contracting Center of Excellence under Contract No. W91WAW-09-C-0134. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Contracting Center of excellence or the official policy or position of the Department of Defense or the US Army Research Institute.

REFERENCES

Campbell, R.C., Keenan, P.A., Moriarty, K.O., Knapp, D.J., & Heffner, T.S. (2004). The Army PerformM21 Demonstration Competency Assessment Program Development Report (Technical Report 1152). Arlington, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Fowlkes, J. E., & Burke, C.S. (2005). Event-based approach to training (EBAT). In N. Stanton, Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H. (Ed.), *Handbook of human factors and ergonomics methods* (pp. 47-41 - 47-45). Boca Raton, FL: CRC Press.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory (3rd ed.).* New York: McGraw Hill, Inc.

Ramachandran, S., J. Ludwig, E. Salas, M. Rosen (2008) A Methodology for Simulation-based Job Performance Assessment. *Proceedings of the Industry/Interservice, Training, Simulation & Education Conference (I/ITSEC 2008)*.

Sackett, P. R., & Laczo, R. M. (2003). Job and work analysis. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology* (Vol. 12, pp. 21-37). Hoboken, NJ: John Wiely & Sons, Inc.

Schank, R. (1995), *What We Learn When We Learn by Doing*, Technical Report no. 60, Institute of Learning Sciences, Illinois.