Incremental Validity of Interactive Multimedia Simulations in Two Organizations

Chris D. Fluckinger*, Nicole M. Dudley and Marisa Seeds*****

*Department of Natural and Social Sciences, Firelands College, Bowling Green State University, 1 University Drive, Huron, OH 44839, USA. cdfluck@bgsu.edu
**Shaker Consulting Group, Cleveland, OH 44122, USA
***Shaker Consulting Group, Cleveland, OH 44122, USA

Interactive multimedia simulations are conceptually distinct from other simulations commonly used as selection tools, such as assessment centers and situational judgment tests, and represent a potentially cost-effective and job-related candidate assessment tool. This study investigated the predictive validity of these simulations with a combined sample of call center employees from two organizations. Results indicate that customized interactive multimedia simulations demonstrate substantial criterion-related validity and significant incremental validity over other noncognitive measures such as biodata and personality.

1. Introduction

Simulations have been recommended for a variety of purposes, including providing recruits with job-specific information (Shotland, Alliger, & Sales, 1998) and collecting information from applicants to make employment selection decisions (Roth, Bobko, & McFarland, 2005). Currently, little is known about the criterion-related validity of interactive multimedia job simulations, although they are often theorized to predict job performance (Ployhart, Schneider, & Schmitt, 2006). The current study investigates the criterion-related validity of interactive multimedia simulations.

Although the term *simulation* may seem intuitive, there is some overlap among researchers and practitioners between the terms *job simulation* and *work sample*. For example, paper-and-pencil situational judgment tests are often referred to as low-fidelity simulations (Motowidlo, Dunnette, & Carter, 1990), and various interactive exercises in assessment centers have been labeled as simulations (O'Connell, Hattrup, Doverspike, & Cober, 2002). These seemingly diverse assessment methods fall under the same general simulation classification largely because they involve situations and activities that real employees would likely encounter (Shotland et al., 1998). This is similar to the description provided by Ployhart et al. (2006) in categorizing work samples as a method in which the applicant performs a set of actual tasks that are similar, both physically and/or psychologically, to those performed on the job. Based on these definitions, it is difficult to distinguish between work samples and job simulations.

One key dimension helps to differentiate work samples and job simulations: fidelity. As Motowidlo et al. (1990) explain, 'fidelity decreases as stimulus materials and responses become less and less exact approximations of actual job stimuli and responses' (p. 640). Work samples are thought to represent high-fidelity assessment methods, and simulations are thought to represent low- to moderate-fidelity assessments (Roth et al., 2005, but see Motowidlo et al., 1990). In a meta-analysis of work samples, Roth et al. (2005) explicitly did not include simulations (such as assessment centers, situational judgment tests, or telephone role-plays). Thus, simulations can be generally classified as low- to moderate-fidelity assessment methods that share conceptual overlap with work samples.

Interactive multimedia simulations are distinct from other simulation methods used for employee selection because they are explicitly designed to be delivered via computer (multimedia) and can be updated in real time based on how the user chooses to complete the exercise (interactive). For example, in the interactive multimedia simulation presented in this study, job applicants must actively monitor and reset a call queue while simultaneously verifying data with a simulated customer call.

International Journal of SELECTION AND ASSESSMENT

In contrast, situational judgment tests are often delivered in video (multimedia) format (Chan & Schmitt, 1997), but the actual content and experience is identical for each participant and thus is not interactive. On the other hand, assessment centers are interactive (i.e., role-playing with a real person) but often not multimedia (O'Connell et al., 2002). We are not aware of any studies directly investigating the criterion-related validity of an interactive multimedia simulation; however, given the conceptual overlap with historically valid methods such as work samples, we expected simulation performance to correlate with overall job performance (Hypothesis 1). To further strengthen inferences regarding the utility of interactive multimedia simulations as selection tools, it is important to show evidence of incremental validity over established selection methods. Provided that an interactive multimedia simulation is designed to measure key competencies for a particular job or job family, the simulations should capture unique variance in job performance not accounted for by other noncognitive assessment methods (Hypothesis 2).

2. Method

2.1. Participants and procedure

The sample consisted of incumbent call center employees in two organizations, and the data were collected as part of concurrent validation studies. Overall, 183 employees in a sales call center (127 male, 56 female; 117 White, 27 Hispanic/Latino, 25 Black, 14 Asian; 121 are under age 40, and 62 are 40 and over) and 200 in a customer service call center (153 male, 47 female; 127 White, 22 Hispanic/Latino, 46 Black, 10 Asian; 138 are under age 40, and 62 are 40 and over) completed the assessment battery. Performance ratings were obtained directly from supervisors. For the sales call center, 200 employees representative of that position within the organization were invited to participate in the validation study. Completing the assessment battery were 191 incumbents, and performance ratings were collected for 183 incumbents from 12 supervisors. For the customer service call center, 200 incumbents completed the assessment battery (out of 240 invited), and performance ratings were collected for all participants from 35 supervisors. The selection content was identical for both organizations except for customized material in the interactive multimedia simulations. Many leading-edge security safeguards were implemented (i.e., eliminating 'print screen', preventing toggling between the assessment window and other programs) to reflect best practice in the use of unproctored, Internet-based testing. Given the similar demographics, study procedures, and assessment batteries of the organizations, we combined the samples in all subsequent analyses. The combined sample of 378 includes only individuals with completed assessment data and performance ratings.

2.2. Measures

2.2.1. Personality

Three personality scales were chosen to include based on their theoretical and empirical relationships with dimensions of job performance critical to both organizations (Allworth & Hesketh, 2000; Grandey, 2000): self-efficacy (five items), locus of control (five items), and work ethic (five items). Personality scales were forced-choice measures on 4-point scales, where the incumbent had to choose one of two statements that was 'most like him/her'. These scales have been rigorously tested and validated within a number of organizations and call center environments. Alpha values (self-efficacy, $\alpha = .67$; locus of control, $\alpha = .67$; and work ethic, $\alpha = .77$) were lower than traditional Likert-based personality measures but consistent with many forced-choice measurement formats. We note that these measures may differ from traditional personality measures commonly presented in the research literature, but we include them as control variables to test the incremental validity of interactive multimedia simulations over self-report, noncognitive measures.

2.2.2. Biodata

Two biodata scales were included based on previous experience with call center organizations: call center experience (duration and breadth of call center work history) and call center success (self-reported previous strong performance in call centers). These scales are similar to those used in past studies for predicting call center performance using both verifiable and nonverifiable scales (Harold, McFarland, & Weekley, 2006). The call center experience scale consists of four verifiable items, including how many years of work experience in a call center or selling over the phone an individual possesses (7-point scale with anchors from *none* to *10 + years*). The call center success scale consists of four less- or nonverifiable items, including the individual's perception of how his/her previous employer would rate his/her call center performance (6-point scale with anchors from *one of the best* to *needing much improvement*). The alpha values (call center experience, $\alpha = .67$; call center success, $\alpha = .65$) were low from a traditional psychometric standpoint but generally consistent with studies utilizing biodata designed from a construct, as opposed to empirical, perspective (Allworth & Hesketh, 2000).

2.2.3. Call center simulation

This exercise was designed to simulate actions that call center employees split their attention between. It contains four simulated calls between a customer and a call center representative. Each simulated call is used to complete two different tasks: (1) entering and verifying customer information, and (2) monitoring a call queue. For each simulated call, the computer screen is divided into two sections: a data entry form with various customer information fields and a dynamically generated call queue monitor. While the call queue appears random to the participant, each call queue is the same. Entering and verifying data is done by listening to the prerecorded simulated interaction between a customer and a call center representative. While listening, the test-taker verifies or enters customer data that mentioned on the call. At the same time, the interactive call queue is updating. The call queue displays two simulated components: average caller wait time and the total number of callers on hold. If either exceeds the threshold (e.g., if the average caller wait time exceeds 2 min), the test-taker must click a button to reset the call queue, which restarts the call queue cycle.

Three measures were captured from each simulated call and used to ultimately to create a simulation composite score: time spent per call, data entry/verification accuracy, and latency (i.e., the time elapsed before the person clicked the button upon a threshold being exceeded). Sum scales for each component were formed by totaling scores across the four simulated calls (e.g., total accuracy, total time spent, etc.). Because the component scores were on different scales, standardized component scores were created and capped at ±3 standard deviations to account for outliers. The simulation component alpha values were acceptable: time per call $\alpha = .77$, latency $\alpha = .94$, and accuracy $\alpha = .82$. These components were combined into one standardized simulation composite score.

2.2.4. Performance ratings

Three performance dimensions, based on job analysis and competency model overlap between both organizations, are included: problem solving, work quality, and customer service. Ratings were based on a 7-point scale (*among the worst* and *among the best*). Alpha values (problem solving, $\alpha = .93$; work quality, $\alpha = .87$; and customer service, $\alpha = .86$) were acceptable. Given the high correlations between these dimensions, we averaged the three to create an overall job performance variable for subsequent analyses.

3. Results

Correlations, means, and standard deviations of all data are presented in Table 1. Hypothesis 1 stated that simulation performance would be positively related to job performance ratings and is supported ($r = .24$, $p < .01$). Hypothesis 2 stated that interactive multimedia simula-

Table 1. Means, standard deviations and correlations

Variable	Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	36.32	11.05												
2. Gender[a]	1.55	.50	.02											
3. Race[b]	1.49	.50	−.07	.22**										
4. Self-efficacy	3.31	.50	−.05	−.17**	−.09									
5. Locus of control	3.22	.59	−.05	−.07	−.07	.48**								
6. Work ethic	2.61	.77	−.05	.07	−.04	.26**	.19**							
7. Call center experience	3.34	1.41	.22**	−.17**	−.10	.15**	.12*	.12*						
8. Call center success	4.71	.75	.10	−.12*	−.04	.38**	.27**	.28**	.32**					
9. Call simulation [c]	0	1	−.32**	.03	.09	.09	−.02	.07	−.19**	−.05				
10. Problem solving	4.80	1.17	−.14**	.04	.06	.16**	.04	.19**	−.05	.19**	.24**			
11. Work quality	4.94	1.04	−.12*	.15**	.03	.03	.03	.18**	−.15**	.13*	.22**	.73**		
12. Customer service	4.89	1.07	−.07	.03	.05	.18**	.14**	.19**	−.08	.21**	.19**	.74**	.74**	
13. Overall performance	4.87	.99	−.12*	.08	.05	.14**	.08	.20**	−.10	.20**	.24**	.91**	.90**	.90**

Note: $N = 378$. [a]Coded 1 = male, 2 = female; [b]coded 1 = white, 2 = nonwhite; [c]presented as standardized scores.
*$p < .05$, **$p < .01$.

tion performance would predict performance ratings incrementally. A hierarchical regression was performed with overall job performance as the dependent variable. The personality and biodata scales were entered in Step 1, and simulation performance was entered in Step 2. Table 2 presents this analysis, and the significant increase in variance accounted for (ΔR^2F) in Step 2 provides support for Hypothesis 2. Self-reported work ethic and past call center success positively predict job performance ratings, and the interactive multimedia simulation accounts for additional variance in job performance beyond the other predictors. Additional regressions with each of the three performance dimensions as dependent variables show significant variance accounted for in Step 2, providing further support for the incremental validity hypothesis.

To investigate whether the multimedia simulations and other predictors demonstrated prediction bias for key demographic categories, we conducted a series of moderated regressions, presented in Table 3. For gender (male/female), race (white/nonwhite) and age (under 40/40 and over), the dichotomous demographic variable was entered into a model with job performance ratings as the dependent variable (the first row of Table 3). Next, for each assessment predictor, the demographic and predictor were entered in Step 1 (not presented in Table 3), and the interaction term was entered in Step 2, with ΔR^2F indicating the incremental significance of the interaction term over the Step 1 model. Of the 18 moderated regressions, only one of the interaction terms was significant: self-efficacy (a small effect; $\Delta R^2 = .01$). Taken together, the results provide little evidence of prediction bias, for the interactive multimedia simulation or the other assessment battery components.

Table 2. Incremental validity analysis

Model	β	R^2	ΔR^2	ΔR^2F
DV: Overall performance				
Step 1		.10		
Self-efficacy	.11			
Locus of control	−.01			
Work ethic	.20**			
Call center experience	−.13			
Call center success	.26**			
Step 2		.14	.04	18.47**
Call simulation	.21**			

Note: $N = 378$.
*$p < .05$, **$p < .01$.

4. Discussion

This investigation attempted to first distinguish interactive multimedia simulations from similar tools and terms often encountered in the literature, such as work samples, situational judgment tests, and assessment centers. Although there is considerable research involving simulations at various levels of fidelity and interactivity, there is little research specifically targeting interactive multimedia simulations. Criterion-related validity evidence detailed in this study shows that interactive multimedia simulations demonstrate both moderate correlations with job performance ratings and predict performance incrementally over other noncognitive selection tools. These findings provide initial evidence that interactive multimedia simulations can offer utility for selection practitioners as job-related selection tools.

Although the results of this study are generally supportive of interactive multimedia simulation criterion-related validity, the generalizability of the findings could be limited by the inclusion of call center positions only. Future research is certainly needed to determine the boundary conditions of interactive multimedia simulation validity across different kinds of jobs, and with predictive validation study designs. Research is also needed to identify challenges in creating job-related simulation content across various job types.

From a psychometric standpoint, there is still considerable debate regarding the construct validity of assessment methods, particularly regarding more intuitively job-related methods such as assessment centers, situational judgment tests, and simulations. Future research replicating the incremental validity findings of this study

Table 3. Moderated regressions for assessments and demographics in predicting overall performance ratings

Variable	Gender[a]				Race[b]				Age[c]			
	β	R^2	ΔR^2	ΔR^2F	β	R^2	ΔR^2	ΔR^2F	β	R^2	ΔR^2	ΔR^2F
Demographic	.08	.01			.05	.00			−.15	.02		
× Self-efficacy	−.44	.03	.00	1.49	−.63	.08	.01	2.82	−.73*	.05	.01	4.09*
× Locus of control	.25	.02	.00	.60	.27	.01	.00	.74	−.61	.04	.01	3.72
× Work ethic	.16	.05	.00	.45	−.27	.04	.00	1.42	.14	.05	.00	.37
× Call center experience	−.18	.02	.00	.93	−.44	.03	.01	5.13	−.26	.03	.00	1.50
× Call center success	−.02	.05	.00	.00	−.66	.05	.01	3.50	.14	.06	.00	.17
× Call simulation	−.10	.07	.00	.38	.06	.05	.00	.13	.12	.06	.00	.59

Note: $N = 378$. [a]Coded 1 = male, 2 = female; [b]coded 1 = white, 2 = nonwhite; [c]coded 1 = under 40, 2 = 40 and older.
*$p < .05$, **$p < .01$.

using more traditional personality measures (i.e., Big 5) would be welcome, as well as designs that include cognitive ability measures. Future research is also needed to investigate the usefulness of interactive multimedia simulations beyond criterion-related validity, including factors such as assessment reactions and/or perceptions of fairness. Research suggests that highly job-related assessment methods such as simulations may result in more favorable reactions, but studies specifically targeting interactive multimedia simulations are needed.

Given the continuing need to deploy assessments that are job related, valid, cost-effective, and can reach a large volume of applicants, interactive multimedia simulations represent a promising alternative. The results presented here indicate that simulations that are customized for a given job or job family can produce substantial criterion-related validity, as well as predict incrementally over traditional selection tools such as biodata and personality.

References

Allworth, E., & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment*, 8, 137–147.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143–159.

Grandey, A. (2000). Emotion regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, 5, 95–110.

Harold, C. M., McFarland, L. A., & Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment*, 14, 336–346.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640–647.

O'Connell, M. S., Hattrup, K., Doverspike, D., & Cober, A. (2002). The validity of 'mini' simulations for Mexican retail salespeople. *Journal of Business and Psychology*, 16, 593–597.

Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory* (3rd ed.) Mahwah, NJ: Erlbaum.

Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, 58, 1009–1037.

Shotland, A., Alliger, G. M., & Sales, T. (1998). Face validity in the context of personnel selection: A multimedia approach. *International Journal of Selection and Assessment*, 6, 124–130.