

## Video-Based Versus Written Situational Judgment Tests: A Comparison in Terms of Predictive Validity

Filip Lievens  
Ghent University

Paul R. Sackett  
University of Minnesota, Twin Cities Campus

In this study, the authors examined whether video-based situational judgment tests (SJTs) have higher predictive validity than written SJTs (keeping verbal content constant). The samples consisted of 1,159 students who completed a video-based version of an SJT and 1,750 students who completed the same SJT in a written format. The study was conducted in a high stakes testing context. The video-based version of an interpersonally oriented SJT had a lower correlation with cognitive ability than did the written version. It also had higher predictive and incremental validity for predicting interpersonally oriented criteria than did the written version. In this high stakes context, applicants also reacted relatively favorably to the SJTs, although there was no significant difference in face validity between the formats. These findings suggest that SJT format changes be made with caution and that validation evidence is needed when changes are proposed.

*Keywords:* situational judgment test, video, predictive validity

By using motion pictures with sound, it is possible to present test situations that appear more like the real-life situation than can be done with any type of printed material. This seems particularly advantageous in proficiency tests, in which every increase in resemblance between the test and the actual job situation contributes to relevance of the test as a measure of on-the-job performance. (Thorndike, 1949, p. 42)

Undoubtedly, Thorndike's (1949) prediction has become reality as video-based testing has become increasingly popular. Thorndike also rightly pointed out that video technology might be especially advantageous for selection procedures that capitalize on their close resemblance with the criterion. It is therefore not surprising that video technology (and more recently CD and DVD technology) has been successfully applied to situational judgment tests (SJTs; Olson-Buchanan & Drasgow, 2006; McHenry & Schmitt, 1994; Weekley & Jones, 1997). One advantage of video-based SJTs is that the increased fidelity of presenting the situations in video format might enhance the point-to-point correspondence to the criterion, leading to higher predictive validity. Another advantage is that SJTs' higher realism might result in more favorable applicant reactions. Finally, video-based SJTs might have lower corre-

lations with cognitive ability, leading to less adverse impact. Downsides of video-based SJTs include the increased costs and equipment requirements.

Although these advantages and disadvantages are often put forward in the context of video-based SJTs, only a couple of studies have systematically investigated the effects of the presentation format (video based vs. written) of SJTs. Chan and Schmitt (1997) conducted a laboratory experiment, showing that a video-based SJT had significantly less adverse impact than a written SJT. In addition, students perceived the video-based SJT to be more face valid than the written SJT. In another laboratory study, Richman-Hirsch, Olson-Buchanan, and Drasgow (2000) found that students reacted more favorably to a multimedia format of a conflict resolution skills SJT as compared with a written format of the same test.

Despite these interesting findings, many questions regarding video-based SJTs remain unanswered. No studies have examined whether video-based SJTs might indeed have higher predictive validity than written SJTs (holding verbal content constant). This is an important question as the utility of any selection procedure depends on the validity of the procedure. If video-based SJTs have no higher validity than their written counterparts, an important argument in defense of their higher developmental and administration costs is negated. Therefore, this study's central research question is as follows: Does an interpersonally oriented video-based SJT lead to better predictions of conceptually linked criteria than a written one? Apart from predictive validity, we also examined the impact of SJT presentation format (video based vs. written) on other conceptually relevant variables such as correlation with cognitively oriented predictors, incremental validity, and applicant perceptions.

The SJT under investigation was used in an actual selection context (in this case high stakes testing), increasing the external validity of our findings. In addition, selection decisions were made

---

Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Ghent, Belgium; Paul R. Sackett, Department of Psychology, University of Minnesota, Twin Cities Campus.

A previous version of this article was presented in April 2005 at the Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles. We acknowledge Steve Motowidlo for his valuable suggestions on an earlier version of this article.

Correspondence concerning this article should be addressed to Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000, Ghent, Belgium. E-mail: Filip.Lievens@ugent.be

in part on the basis of SJT scores, enabling us to use a predictive validation design. In particular, the SJTs under study were administered as part of the admission exam for medical and dental studies in Belgium. SJTs had been administered in a video-based format from 1997 to 2002. Yet, in 2003, cost and technological concerns prompted the governmental body overseeing the examination to change the format from video based to written.

### Development of Hypotheses

One of the characteristics of SJT items in general is that they are ambiguous because they present practical contextualized problems to applicants that are ill-defined and incomplete in information and have multiple possible solutions (Schmitt & Chan, 2006; Hedlund et al., 2001). SJT items in general, including the ones used in this study, also often represent interpersonal situations (Weekley, Ployhart, & Holtz, 2006). Probably, a written medium is not able to match the ambiguity level inherent in SJT items because a written medium cannot convey the multiplicity of cues (body language, voice tone, and inflection) present in interpersonal interactions. Conversely, the video-based administration medium provides a much richer assessment environment. Along these lines, Olson-Buchanan and Drasgow (2006) posited the following:

[In video-based SJTs assessee] see and hear people interacting, perceive or, importantly, fail to perceive their emotions and stress, and confront dilemmas about one's choice of action or inaction . . . With this format, we may be able to better understand how the assessee will interpret verbal and nonverbal behaviors of others in the workplace and choose to respond. (p. 253)

In other words, the richer and more detailed portrayal of SJT items through video might lead to more fidelity and a better match with criterion behavior as presented on the job. In turn, this might result in higher criterion-related validity.

Although it makes sense to expect that a video-based format might have higher fidelity and therefore higher predictive validity than a written format, the opposite has also been argued. It has been mentioned that video-based SJTs might insert irrelevant contextual information and unintentionally bring more error into SJTs (Weekley & Jones, 1997). In turn, this might result in lower predictive validity. So far, there is little empirical support for this view. Various studies (reviewed by Olson-Buchanan & Drasgow, 2006) have confirmed the good criterion-related validity of video-based tests. Yet, as noted by several authors (e.g., Olson-Buchanan & Drasgow, 2006; Weekley & Jones, 1997), no research has examined whether the validity of video-based SJTs is indeed higher than that of their written counterparts (holding SJT verbal content constant). On the basis of the arguments mentioned above, we expected video-based SJTs to have higher validity for conceptually relevant criteria than written SJTs. Applied to this study, this leads to the following hypothesis.

*Hypothesis 1:* An interpersonally oriented video-based SJT will have higher validity for predicting interpersonally oriented criterion performance than will a written SJT (holding verbal content constant).

Apart from predictive validity, the relationship of SJTs with cognitive ability might also differ for a video-based versus a

written administration. In their meta-analysis, McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001) found a wide variation in the correlation between cognitive ability and SJTs. The presentation format of SJTs might explain some of this variability. In particular, the greater reading component inherent in a written administration might lead to a stronger correlation between cognitive ability and the SJT (see Goldstein, Yusko, Braverman, Smith, & Chung, 1998). Indirect support for this expectation comes from Chan and Schmitt's (1997) finding of video-based SJTs having lower adverse impact than written SJTs. If video-based SJTs have a lower correlation with cognitive ability, video-based SJTs should also have more incremental validity over cognitively oriented predictors for predicting relevant criterion performance than written SJTs. This leads to the following hypotheses.

*Hypothesis 2a:* The correlation between an interpersonally oriented SJT and cognitively oriented predictors will be significantly higher when the SJT is administered in a written format as compared with a video-based format.

*Hypothesis 2b:* An interpersonally oriented video-based SJT will have more incremental validity over cognitively oriented predictors in predicting interpersonally oriented criterion performance than will a written SJT.

Finally, the SJT presentation format might also affect applicant perceptions. Specifically, we expect the presentation format to influence a procedural justice dimension such as job relatedness (face validity) because this dimension refers to the fidelity of the test in representing criterion behavior. Prior studies have already tested this expectation in a research (laboratory) context. Chan and Schmitt (1997) found that students' face validity perceptions were significantly higher for a video-based SJT than for a written SJT. In addition, Richman-Hirsch et al. (2000) demonstrated that a multimedia SJT was seen as more face valid, more enjoyable, and more modern than the computerized and written forms of the same SJT. No studies, however, have compared applicant perceptions of video-based versus written SJTs in an actual selection context. This gap in the literature is important because in actual selection the stakes are much higher than in a research context. On the basis of prior laboratory studies, we formulated the following hypothesis.

*Hypothesis 3:* Applicants' face validity perceptions of a video-based SJT will be significantly higher than their perceptions of the same SJT in written form.

## Method

### Sample

As noted, SJTs had been administered in a video-based format in medical admission exams in Belgium from 1997 to 2002. In 2003, cost and technological concerns prompted the governmental body overseeing the examination to change the SJT presentation format from video-based to written. To enable an investigation of the effect of presentation format, we transformed the video-based SJT used in 2000 to a written format that was used as the SJT in the admission exam of 2003.

In 2000, 1,159 students (65.2% women and 34.8% men; 99.5% White; mean age = 18.8 years) attended the medical admission exam in Belgium and completed the video-based version of the SJT. In 2003, 1,750 students (65.2% women and 34.8% men; 99.5% White; mean age = 18.11 years) attended the admission exam and completed the written version of the SJT. As can be seen, the demographic characteristics of these samples are virtually the same. This is not surprising because these are essentially populations (i.e., all students who participated in a specific year in the admission exam). No students participated in both exams (2000 and 2003).

### *Development of the SJT*

Similar to Chan and Schmitt (1997), the original SJT was a video-based SJT. This SJT aimed to measure interpersonal and communication skills. We used an approach analogous to other studies (see, e.g., Weekley & Jones, 1997) for developing the SJT. First, we collected realistic critical incidents regarding interactions between physicians and patients from experienced physicians and professors in general medicine. For example, incidents dealt with handling complaints of a patient or with conveying bad news. Second, vignettes that nested the critical interpersonal incidents were written. Two professors teaching physicians' consulting practices tested these vignettes for realism. Similarly, questions and response options were derived. For test security reasons, pilot testing and calibration of these questions were not possible. Third, semiprofessional actors were hired and videotaped in a recording studio. To guarantee realism, an experienced physician attended the set. Finally, a panel of experts (experienced physicians, professors in general medicine) was asked to develop a scoring key. Agreement among the experts was generally satisfactory (Cohen's  $\kappa$ s > .70) and discrepancies were resolved upon discussion, leading to the scoring rule. This scoring key indicated which response alternative was correct for a given item (+1 points). It was forbidden by law to use different scoring rules (e.g., penalizing for choosing an incorrect alternative by assigning -1 points). In its final form, the video-based SJT consisted of short videotaped vignettes of key interpersonal situations that physicians are likely to encounter with patients. A narrator introduced each vignette. Per situation, the scene froze at an important point and candidate medical students had 25 s to answer the question related to the scene presented. In total, the SJT consisted of 30 questions of the multiple-choice type, with four response alternatives.

Next, we developed a written version of the video-based SJT. In creating written versions of existing video-based SJTs, there are two major options. Both options differ from the video SJT but in different ways. Both options hold verbal content constant. One option also attempts a verbal description of nonverbal aspects of the video (e.g., describing facial expressions, tone of voice), whereas the other option does not. Adding a verbal description of nonverbal aspects creates a key difference between the video-based and the written SJT versions in that in the video-based SJT participants are required to detect the presence of and make inferences about nonverbal features; these detection and inference processes are not required when these nonverbal features are described in writing. On the other hand, not including information about nonverbal features means that some information in the video SJT is missing from the written SJT. In short, both methods of producing a written SJT differ from the video SJT in terms of available information and candidate information processing requirements. Ideally, one might conduct an experiment contrasting both options with the video SJT. However, the current study took place in an operational testing context; because scores were used to rank candidates, it was crucial that all were treated in the same manner (i.e., received the same test). Thus, we were forced to choose one of the two options, and for the current study, we opted to limit the written SJT to describing verbal content (see also Chan & Schmitt, 1997). Specifically, each of the vignettes was transcribed in dialogue form so that the verbal content of each written situation was identical to that of the respective videotaped situation. The answer booklet

that the participants used was the same for the video-based SJT and the written SJT. The scoring key was also the same in both SJTs. The testing time of both SJTs was 45 min.

The internal consistency coefficients of the SJTs across administration formats were very similar ( $\alpha = .31$  for video-based SJT;  $\alpha = .29$  for written SJT). SJTs typically demonstrate low internal consistency because the situations and response options presented by SJTs are often multidimensional (Chan & Schmitt, 1997, 2002; Clause, Mullins, Nee, Pulakos, & Schmitt, 1998).

### *Other Measures*

Apart from the SJT, the admission exam consisted of a cognitive section and a work sample. The cognitive section was composed of four science knowledge tests (chemistry, physics, mathematics, and biology) and a cognitive ability test. The cognitive ability test contained 50 items with five response alternatives. Each year, these items were randomly selected from a larger item pool. The items were formulated in verbal, numeric, or figural terms. Hence, this was a broad cognitive ability test that aimed to measure general mental ability. An example numeric item of the cognitive ability test is as follows: "Complete the following series of numbers: 54 81 18 27 6 9 ?" In light of test security, we cannot mention the source of these cognitively oriented tests. There were relatively high intercorrelations among the cognitively oriented tests (mean  $r = .36$ , range = .29-.45). Hence, we computed a cognitive composite to use in our analyses. This is in line with results from a recent meta-analysis (Kuncel, Hezlett, & Ones, 2001) that showed that a composite of general measures (e.g., Graduate Record Exam [GRE] verbal and numerical) combined with specific GRE subject-matter tests provided the highest validity in predicting academic performance. The internal consistencies of our five-test cognitive composite were .74 (2000) and .71 (2003).

The work sample was a miniaturized sample of tasks that students would encounter in their medical education (i.e., reading and understanding an article with a medical subject matter). The 10-page text was developed on the basis of a medical text in a popular medical journal or handbook. In 2000, the text dealt with headaches, whereas the 2003 text was about diabetes. Students had 50 min to read the text and to answer 30 multiple-choice questions (with four response alternatives).

To make actual admission decisions, we computed a weighted sum of each of the predictors (cognitively oriented tests, work sample, and SJT). Next, a minimal cutoff was determined on this operational composite. The weights and cutoff score were determined by law. In both 2000 and 2003, the passing rate of the exam was about 30%. Candidates who passed the exam received a certificate that warranted entry in any medical university. There was no further selection on the part of universities. However, not all students passing the exam chose to enter medical school.

### *Face Validity Measure*

At the end of the admission exam, candidate medical students were invited to fill out a short questionnaire about their perceptions of the admission exam. To guarantee honesty in responding, we emphasized that completion of this inventory was voluntary and anonymous so that responses to this inventory could not be linked to the admission exam scores. In 2000, the response rate to this questionnaire was 55%; in 2003 the response rate was 61%. Among other questions, this inventory measured students' face validity perceptions of the various admission exam parts with four items adapted from Smither, Reilly, Millsap, Pearlman, and Stoffey (1993). An example item was "The actual content of the SJT was clearly related to medical education." Items were rated on a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items were averaged to compute an overall face validity rating per admission exam part. The internal consistencies of this measure were .76 (for measuring

face validity perceptions of the video-based SJT in 2000) and .66 (for measuring face validity perceptions of the written SJT in 2003).

### Criterion

Criterion data were gathered from students who had passed the exam and had completed the 1st year of medical studies in one of the universities in Belgium. As a first criterion measure, we retrieved archival data on students' grade point average (GPA) from all universities. GPA serves as a more cognitively oriented criterion because it mainly consists of science and medical-related subjects in the first years of medical school. Given differences across universities, we standardized students' GPA within university<sup>1</sup> and academic year.

As a second criterion measure, we retrieved archival data on students' scores on interpersonally oriented courses.<sup>2</sup> In these courses, interpersonal and communication skills are taught. Hence, these courses typically consist of interactional exercises and exams. Note that these courses are not void of medical subject matter. However, the medical subject content is secondary to the skills of communicating with patients, asking questions, dealing with their complaints, and so forth. We gathered interpersonally oriented criterion data because they are especially useful for validating the SJT used in the admission exam. Recall that this SJT aimed to measure interpersonal and communication skills.

To ensure that the criterion had not changed from 2000 to 2003, we included only scores on interpersonally oriented courses that had exactly the same content description and were taught by the same professors. Further evidence that the criterion had not changed is given by the fact that the correlation between the interpersonally oriented criterion and GPA was similar in 2000 and 2003 (.55 and .61, respectively). Given that interpersonal courses were not included in the curriculum of some universities, interpersonally oriented criterion data were based on a smaller sample size than was GPA in both 2000 and 2003.

### Results

Tables 1 and 2 present the means, standard deviations, and correlations among the study variables in the two samples (2000 and 2003). As can be seen, the difference between the mean score on the video-based SJT ( $M = 15.86$ ,  $SD = 2.45$ ) and the written SJT ( $M = 15.68$ ,  $SD = 2.46$ ) was small,  $d = .07$ ,  $t(2,627) = 1.65$ ,  $ns$ . In addition, the difficulty indices for the items across the two administration modes correlated .97. We note that the two other predictors do show mean differences between 2000 and 2003. For the medical text, this pattern is not unexpected, as the difficulty of the text has been found to vary across years. For the cognitive composite the lower scores in 2003 may reflect either a change in the composition of the applicant pool or a change in the difficulty of the science tests.<sup>3</sup>

Hypothesis 1 dealt with the effects of presentation format on predictive validity. As noted above, the admission decision was made on the basis of a third variable (the operational composite). Only students who successfully passed the admission exam (i.e., scored higher than the cutoff determined on the operational composite) undertook medical studies, resulting in indirect range restriction. Given that indirect range restriction is a special case of multivariate range restriction, we applied the multivariate range restriction formulas of Ree, Carretta, Earles, and Albert (1994) to the uncorrected correlation matrix. Statistical significance was determined prior to correcting the correlations (Sackett & Yang, 2000). A comparison of the last rows of Tables 1 and 2 shows that the correlation between the video-based SJT and the interpersonal

criterion ( $r = .34$ , corrected  $r = .35$ ) was significantly ( $p < .01$ ) higher than the correlation between the written SJT and the interpersonal criterion ( $r = .08$ , corrected  $r = .09$ ). We also conducted a moderated regression analysis in which data across the presentation formats were pooled and the presentation format and the interaction between presentation format and SJT were added as predictors. This analysis revealed that the interaction of presentation format and SJT was significant ( $p < .05$ ) and explained 1.4% of additional variance, lending support to Hypothesis 1.

The intercorrelations of the SJT with the cognitive composite are the basis for testing Hypothesis 2a. A comparison between Tables 1 and 2 shows that the correlation (.18) between the written SJT and the cognitive composite was significantly ( $p < .05$ ) higher than the correlation (.11) between the video-based SJT and the cognitive composite.<sup>4</sup> This lends support to Hypothesis 2a.

Hypothesis 2b posited that a video-based SJT would have more incremental validity over cognitively oriented predictors in predicting relevant criterion performance than a written SJT. To test this hypothesis, we entered the cognitive test composite as a first block because these tests have been traditionally used in medical admission exams. Next, we entered the medical text in the regression equation. Finally, we entered the SJT. The matrices that we corrected for multivariate range restriction served as input for these regression analyses. Statistical significance was determined prior to correcting the correlations (by conducting the same regression analyses on the uncorrected matrices). A comparison between Tables 3 and 4 yields some interesting findings. The video-based SJT accounted for a significant and substantial amount of variance in the interpersonal criterion (11%,  $p < .01$ ). Conversely, the written SJT did not account for incremental variance in the interpersonal criterion. The written SJT did explain a

<sup>1</sup> Within-school standardization might not be appropriate if students of differing ability went to different schools. To examine this possibility, we followed the same procedure as Cullen, Hardison, and Sackett (2004). We first standardized the GPA data within each university, as the grading scales were not constant from school to school. We then regressed GPA on the total admission score (i.e., operational composite) and a set of dummy variables representing each university. Results showed that none of the coefficients for the dummy variables were significant, indicating that university choice was not related to the total admission score.

<sup>2</sup> Apart from these two criteria, we also computed a noninterpersonal GPA criterion. Results for noninterpersonal GPA were the same as for overall GPA. This is not surprising as in the first years of medical education overall GPA is mainly determined by noninterpersonal (medical and science related) aspects.

<sup>3</sup> Probably, both factors applied here. On the one hand, the science knowledge tests were somewhat more difficult in 2003. On the other hand, the applicant pool in 2003 consisted of a larger group of students from the Netherlands than in 2000 (19.2% vs. 12.9%, respectively). Typically, these Dutch students scored lower on the cognitively oriented tests. Therefore, we also ran our analyses without this group of students from the Netherlands. Results were the same. Note also that in the hierarchical regression analyses the cognitively oriented tests are entered in a first step so that their effects on validity are taken into account.

<sup>4</sup> Similar patterns could be observed for the correlation between the SJTs and the cognitive ability test alone ( $r = .13$  for the video-based SJT;  $r = .20$  for the written SJT) and for the correlation between the SJTs and the written text ( $r = .20$  for the video-based SJT;  $r = .24$  for the written SJT).

Table 1  
Descriptive Statistics of Study Variables in 2000 (SJT Administered in Video-Based Format)

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
Predictors ( <i>n</i> = 1,159)								
1. Cognitive composite	15.18	3.73						
2. Medical text	13.55	3.72	.35**					
3. Video-based SJT	15.86	2.45	.11**	.20**				
4. Operational composite	21.59	3.99	.96**	.54**	.28**			
Criteria								
5. GPA ( <i>n</i> = 395)	0.00	1.00	.32** (.40)	.02 (.10)	.07 (.10)	.31** (.39)		
6. Interpersonal criterion ( <i>n</i> = 145)	0.00	1.00	.10 (.13)	-.04 (-.01)	.34** (.35)	.14 (.17)	.55** (—)	

Note. Corrected correlations are in parentheses. Correlations were corrected for multivariate range restriction. Statistical significance was determined prior to correcting the correlations. A dash indicates that no correction for range restriction was applied to the criterion. SJT = situational judgment test; GPA = grade point average.

\*\*  $p < .01$ .

significant albeit small amount (1%,  $p < .05$ ) of additional variance in GPA.

Hypothesis 3 posited that candidates' face validity perceptions would be higher for the SJT when administered in a video-based format as compared with a written format. As these perceptions were measured at the end of an admission exam (high stakes context) and the response rate was about 60%, they should be interpreted with caution. Although these measures were anonymous, it is still possible that some respondents may have believed that they could somehow be identified. Results showed that face validity perceptions of the SJT administered in the video-based format equaled 3.41 ( $SD = 0.78$ ), whereas perceptions of the SJT administered in the written format equaled 3.44 ( $SD = 0.96$ ). A *t* test showed that this difference was not significant and the effect size was small ( $d = -.03$ ), lending no support to Hypothesis 3. To make a fair comparison between the face validity of the two SJTs, we corrected the results for unreliability in the face validity measure used (on the basis of the internal consistency value). Recall that the internal consistency of the face validity measure in 2000 was .66, whereas it was .76 in 2003. The corrected means are 4.20 for the video-based SJT and 3.94 for the written SJT. Note that the face validity means for both SJTs were much higher than the other parts of the admission exam. For instance, the face validity per-

ceptions of the cognitive ability measure equaled 2.75 (in 2000) and 2.79 (in 2003).

## Discussion

This study addresses the following key question: Can one expect comparable validity findings if a video-based SJT is converted to a written format? The results of our predictive validity study in a field setting answer this question negatively. First, the interpersonally oriented video-based SJT used in this study had significantly higher predictive validity and incremental validity for predicting interpersonally oriented criteria than did the written SJT (holding SJT verbal content constant). Note also that the basis for the validity of this video-based SJT is not its correlation with cognitive ability, as this video-based SJT had a low correlation (.11) with the cognitive composite. Instead, it seems to be fidelity and close resemblance to the criterion that produce good validity. Hence, our study does not support that the video-based format unintentionally introduces irrelevant contextual information and therefore more error into SJTs. This result conforms to a so-called additive model of the impact of different information channels (Archer & Akert, 1980; Gesn & Ickes, 1999). According to this model, adding extra sources of information leads to higher accu-

Table 2  
Descriptive Statistics of Study Variables in 2003 (SJT Administered in Written Format)

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
Predictors ( <i>n</i> = 1,468)								
1. Cognitive composite	13.99	3.65						
2. Medical text	18.70	4.70	.36**					
3. Written SJT	15.68	2.46	.18**	.24**				
4. Operational composite	20.10	4.80	.94**	.60**	.34**			
Criteria								
5. GPA ( <i>n</i> = 657)	0.00	1.00	.33** (.40)	.18** (.25)	.10* (.13)	.36** (.42)		
6. Interpersonal criterion ( <i>n</i> = 269)	0.00	1.00	.03 (.04)	.12 (.12)	.08 (.09)	.09 (.10)	.61** (—)	

Note. Corrected correlations are in parentheses. Correlations were corrected for multivariate range restriction. Statistical significance was determined prior to correcting the correlations. A dash indicates that no correction for range restriction was applied to the criterion. SJT = situational judgment test; GPA = grade point average.

\*  $p < .05$ . \*\*  $p < .01$ .

Table 3  
*Summary of Hierarchical Regression Analyses of Predictors on Grade Point Average (GPA) and Interpersonal Criterion in 2000 (SJT Administered in Video-Based Format)*

Variable	GPA (n = 395)					Interpersonal criterion (n = 145)				
	$\beta$	<i>t</i>	<i>p</i>	$R^2$	$\Delta R^2$	$\beta$	<i>t</i>	<i>p</i>	$R^2$	$\Delta R^2$
Step 1										
Cognitive composite	.40	6.52	.01	.16	.16**	.09	0.94	.35	.02	.02
Step 2										
Medical text	-.03	-0.64	.52	.16	.01	-.07	-0.80	.43	.02	.01
Step 3										
Video-based SJT	.07	1.57	.12	.16	.01	.34	4.25	.01	.13	.11**

*Note.* The corrected matrices served as input for the regression analyses. Statistical significance was determined prior to correcting the correlations (by conducting the same regression analyses on the uncorrected matrices). Parameter estimates are for final step, not entry. Because of rounding,  $\Delta R^2$  differs .01 from the cumulative  $R^2$ . SJT = situational judgment test.  
 \*\*  $p < .01$ .

racy and thus higher fidelity. This additive model has also received support in interview research as interviewees’ visual and aural cues were found to provide valid information over and above the verbal content of their answers (Burnett, Fan, Motowidlo, & DeGroot, 1998; DeGroot & Motowidlo, 1999).

Second, the correlation between cognitively oriented predictors and SJTs was significantly higher for a written SJT than for a video-based SJT. In addition, a written SJT was more predictive of cognitive aspects of the criterion space as measured by GPA. Conversely, a video-based SJT was more predictive of an interpersonal criterion. Apparently, a written version of an SJT places itself somewhat more in the cognitive category and less in the alternative (noncognitive) selection category. The reverse seems to be true for a video-based SJT.

As a third contribution, we examined whether possible differences in terms of face validity perceptions between video-based and written SJTs generalized to a high stakes context. The SJTs used (either video based or written) always received significantly more favorable ratings than the other parts of the exam. Yet, contrary to prior laboratory studies (Chan & Schmitt, 1997; Richman-Hirsch et al., 2000), we did not find significant differ-

ences between the face validity perceptions of the video-based and written SJTs. Apparently, candidates perceived both SJTs to be equally related to activities in medical education. Thus, taken together these results suggest that it makes sense to invest in video-based SJTs for measuring interpersonal skills. Applied to this study, we hope that the governmental body overseeing the admission exam decides to revert back to video-based SJT in the near future, at least until an alternative written format emerges that produces comparable validity.

An important question is whether our finding of higher predictive validity for the video-based SJT generalizes to all SJTs? To answer this question, one should look into the characteristics of the SJTs in this study. First, we focused on an interpersonally oriented SJT. Although most SJTs are interpersonally oriented (Weekley et al., 2006), these types of SJTs might maximize the gap between the video-based versus written presentation format because emotions, facial expressions, and voice inclinations play a predominant role in interpersonal interactions. Future studies should examine whether validity differences between video-based and written SJTs are smaller for SJTs with a less salient interpersonal focus. Second, we transcribed only the verbal dialogue to develop our written SJT.

Table 4  
*Summary of Hierarchical Regression Analyses of Predictors on Grade Point Average (GPA) and Interpersonal Criterion in 2003 (SJT Administered in Written Format)*

Variable	GPA (n = 657)					Interpersonal criterion (n = 269)				
	$\beta$	<i>t</i>	<i>p</i>	$R^2$	$\Delta R^2$	$\beta$	<i>t</i>	<i>p</i>	$R^2$	$\Delta R^2$
Step 1										
Cognitive composite	.35	7.44	.01	.16	.16**	.01	0.09	.93	.01	.01
Step 2										
Medical text	.09	2.28	.02	.17	.01*	.11	1.57	.12	.02	.01
Step 3										
Written SJT	.08	2.04	.04	.18	.01*	.06	0.94	.35	.02	.01

*Note.* The corrected matrices served as input for the regression analyses. Statistical significance was determined prior to correcting the correlations (by conducting the same regression analyses on the uncorrected matrices). Parameter estimates are for final step, not entry. Because of rounding,  $\Delta R^2$  differs .01 from the cumulative  $R^2$ . SJT = situational judgment test.  
 \*  $p < .05$ . \*\*  $p < .01$ .

Nonverbal content was not added. Strictly speaking, this means that our conclusions apply only to written SJTs that do not include information on the nonverbal behavior of the characters.

Future studies might experiment with alternative transcription methods. Apart from the movie script method that we used (the transcription documents what each actor says), an alternative transcription method is more akin to a novel, in which nonverbal and contextual aspects are described in addition to the verbal exchange. Future studies are also needed to systematically compare other SJT response formats (Funke & Schuler, 1998). Contrary to this study's SJT, one might use an audiotape or a proctor to time the respondents' reading of each item and answering period. Accordingly, respondents are told to go to the next item only after a fixed amount of time has passed. Finally, one might try out interactive video-based formats (i.e., formats with branched items in which one's response to the main scene determines which branch scene is shown next). Clearly, we need to examine the criterion-related validity, "fakability," and face validity of these alternative SJT formats.

This study has several possible limitations. It is possible that test security breaches might have confounded our examination. This is because the video SJT of 2000 was transcribed for use in 2003. Therefore, in 2003 students might have been better prepared (either because of coaching or study) for the SJT or might even have known the questions beforehand. Although the field setting of our study precludes us from ruling out this possibility, we believe it is highly unlikely for several reasons. First, the 2000 admission exam was not the first time the SJT format was used; video-based SJTs had been used each year since 1997. Thus, even the students watching the 2000 video-based SJT had a good sense of what the SJT was about and what kind of questions were to be expected. Second, the SJT has a relatively low weight in the operational composite (see Tables 1 and 2). Most coaching and individual study efforts focus on the cognitively oriented tests (especially the science knowledge tests). Note also that none of the students of 2000 participated in 2003. So, there were no repeat test takers. Third, we did not notice an increase in the means in the second administration. If the test content had leaked out, the SJT scores on the second administration would have been markedly higher.

Another possible limitation is related to the low internal consistency of the SJTs. This might be because our SJTs consisted of only 30 items with a dichotomous scoring scheme (SJT items were scored with either 0 or 1). SJTs typically have more items and use a different scoring scheme (see Motowidlo, Dunnette, & Carter, 1990). In addition, relatively low internal consistencies seem to be a common finding because SJTs often capture multiple constructs (McDaniel et al., 2001). It is important to note that this limitation does not confound our central comparison because the internal consistencies across the two formats (written vs. video based) were the same.

A final limitation relates to the setting of this study. Our study was conducted in a high stakes context in Belgium. As the sample consisted of 99.5% White candidates, we could examine neither adverse impact nor differential prediction. In this study, GPA also served as criterion measure. GPA is more comparable to using training performance as criterion rather than job performance. Therefore, future research should examine whether our results

generalize to employment settings with job performance as the criterion.

## References

- Archer, D., & Akert, R. M. (1980). The encoding of meaning: A test of three theories of social interaction. *Sociological Inquiry*, *50*, 393–419.
- Burnett, J. R., Fan, C., Motowidlo, S. J., & DeGroot, T. (1998). Interview notes and validity. *Personnel Psychology*, *51*, 375–396.
- 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.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*, 233–254.
- Clause, C. C., Mullins, M. E., Nee, M. T., Pulakos, E. D., & Schmitt, N. (1998). Parallel test form development: A procedure for alternative predictors and an example. *Personnel Psychology*, *51*, 193–208.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability–job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, *89*, 220–230.
- DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*, 986–993.
- Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, *6*, 115–123.
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for emphatic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, *77*, 746–761.
- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology*, *51*, 357–374.
- Hedlund, J., Plamondon, K., Wilt, J., Nebel, K., Ashford, S., & Sternberg, R. J. (2001, April). Practical intelligence for business: Going beyond the GMAT. In J. Cortina (Chair), *Out with the old, in with the new: Looking above and beyond what we know about cognitive predictors*. Symposium presented at the 16th Annual Convention of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 162–181.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*, 730–740.
- McHenry, J. J., & Schmitt, N. (1994). Multimedia testing. In M. G. Rumsey & C. B. Walker (Eds.), *Personnel selection and classification* (pp. 193–232). Hillsdale, NJ: Erlbaum.
- 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.
- Olson-Buchanan, J. B., & Drasgow, F. (2006). Multimedia situational judgment tests: The medium creates the message. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 253–278). San Francisco: Jossey-Bass.
- Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for restriction of range: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, *79*, 298–301.

- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology, 85*, 880–887.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118.
- Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 135–155). San Francisco: Jossey-Bass.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*, 49–76.
- Thorndike, R. L. (1949). *Personnel selection*. New York: Wiley.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25–49.
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 157–182). San Francisco: Jossey-Bass.

Received December 9, 2004

Revision received July 13, 2005

Accepted July 26, 2005 ■

### *Experimental and Clinical Psychopharmacology*<sup>1</sup>

#### Call for Papers: Special Issue on

#### Recent Advances in Psychopharmacology and Mental Health

A Special Issue of *Experimental and Clinical Psychopharmacology* on advances in the psychopharmacological treatment of mood and anxiety disorders is planned for early 2007. Mood disorders afflict an estimated 20.9 million Americans and include major depressive disorders, bipolar disorders, and dysthymic disorders. Anxiety disorders afflict an estimated 40 million Americans and include posttraumatic stress disorder (PTSD), panic disorder, and obsessive–compulsive disorder (OCD). Treatment of these psychiatric disorders is often complicated by comorbid substance abuse disorders.

Authors are invited to submit original research reports for consideration. All papers will be peer reviewed and must be submitted electronically. This Special Issue will also contain some invited review papers on the status of novel pharmacotherapies such as corticotropin-releasing-hormone (CRH1) receptor antagonists, testosterone supplements, and dehydroepiandrosterone (DHEA) for alleviation of depression and anxiety.

The deadline for submission of original research reports is

**December 16, 2006.**

To submit an article, go to

[http://www.jbo.com/jbo3/submissions/dsp\\_jbo.cfm?journal\\_code=pha](http://www.jbo.com/jbo3/submissions/dsp_jbo.cfm?journal_code=pha)

Inquiries about this special issue should be directed to:

[apa-ecp-ed@mclean.harvard.edu](mailto:apa-ecp-ed@mclean.harvard.edu)

<sup>1</sup>*Experimental and Clinical Psychopharmacology* is published by the American Psychological Association and edited by Nancy K. Mello, Ph.D., McLean Hospital and Harvard Medical School