Applied H.R.M. Research, 2003, Volume 8, Number 1, 33-44

Response Instructions and Racial Differences in A Situational Judgment Test

Nhung T. Nguyen
Lamar University

Michael A. McDaniel
Virginia Commonwealth University

This study explored the effect of response instructions in a paper-and-pencil situational judgment test (SJT) on racial differences. Two different response instructions, i.e., knowledge and behavioral tendency, using the exact same test items were examined in a within-subjects design. Results showed that Black-White differences exist in both versions of the test. Further, racial differences were found to be a function of cognitive saturation of the test. The SJT with knowledge instructions was more cognitively loaded than was the same test with behavioral tendency instructions. Implications for researchers and practitioners are discussed in the context of reducing adverse impact in personnel selection.

Situational judgment tests (SJTs) are designed to assess an applicant's judgment regarding a situation encountered in the work place (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). In such tests, respondents are presented with work-related scenarios and asked to identify an appropriate course of action from a given list of possible courses of action. The assumption underlying SJTs is that how an individual performs on a job simulation predicts future job performance (Motowidlo, Dunnette, & Carter, 1990).

Research on SJTs indicates that these tests are useful and popular selection tools (McDaniel et al., 2001). There are two reasons that might explain the increasing popularity of SJTs as an instrument in selecting employees. First, SJTs have been shown to have substantial validity as indicators of job performance (McDaniel et al., 2001). Second, SJTs have been demonstrated to have less race-based adverse impact (Chan & Schmitt, 1997; Motowidlo & Tippins, 1993; Motowidlo et al., 1990; Weekley & Jones, 1997, 1999) than cognitive measures.

However, several problems in SJTs remain unanswered. Although SJTs show substantial criterion-related validity (McDaniel et al., 2001) and reduced racial adverse impact (Chan & Schmitt, 1997; Motowidlo & Tippins, 1993; Weekley & Jones, 1999), the degree of racial differences varies across tests. Most studies on SJTs (e.g., Chan & Schmitt, 1997; Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1997, 1999) reported subgroup differences ranging from .2 to 1.2 standard deviations with Whites scoring higher than Blacks. Most estimates fell in the middle of this range. Although these subgroup differences are smaller in magnitude than what is normally found in cognitive measures (Hunter & Hunter, 1984), they still produce adverse impact under different selection ratios frequently encountered in actual selection situations (Bobko, Roth, & Potosky, 1999; Sackett & Ellingson, 1997). Further, the question of what determines subgroup differences in SJT performance has not been adequately answered. Chan and Schmitt (1997) showed that the video-based SJT produced smaller Black-

White difference than the traditional paper-and-pencil method. However, the high cost involved in producing and administering a video-based test will often make this a less viable option than the traditional paper-and-pencil tests. Although numerous studies have examined the issue of subgroup differences in SJTs, almost no research has examined the function of response instructions in reducing adverse impact in SJTs. To date, only one study (Ployhart & Ehrhart, 2001) examined the influence of response instructions on the validity and reliability of SJTs. They found that different response instructions caused dramatic differences in validity and reliability of SJTs. To the extent that different response instructions influence validity and reliability, differences in adverse impact resulting from different ways used to instruct test takers to respond may be one of the reasons for the wide range of subgroup differences reported in previous research.

The purpose of this research was to explore the impact of response instructions on racial differences in paper-and-pencil situational judgment test performance. Specifically, the question of how different response instructions for the *exact same* test items might lead to different magnitudes of racial differences in test scores was addressed. We first reviewed approaches for reducing adverse impact and followed by how and why different response instructions should produce different magnitudes of racial differences.

Approaches for Reducing Adverse Impact

Reducing racial differences or adverse impact in selection measures has long been a concern among personnel researchers because equal employment opportunity for all is the goal. Review of the literature on personnel selection indicates that researchers have pursued two strategies to reduce adverse impact. The first strategy is to search for *substantive* predictors of job performance (e.g., conscientiousness) that show less ethnic differences than cognitive abilities. Cognitive abilities, although proved to be of substantial validity in predicting job performance (Hunter & Hunter, 1984), have the largest Black-White adverse impact (average d = 1.0) favoring Whites (Roth, BeVier. Bobko, Switzer, & Tyler, 2001). Yet, the overall 1.0 standard deviation difference favoring Whites remains large, causing dramatic race-based adverse impact in selection. (For a review of this strategy in reducing adverse impact, see Bobko et al., 1999).

The second strategy is to search for a *method* that shows the least adverse impact. For example, structured interviewing compared to bio-data has been shown to have less adverse impact (d=.23 vs. d = .33) (Bobko et al., 1999). There are a few studies showing that SJTs evidence less adverse impact than cognitive selection measures. Most studies reported Black-White differences ranging from .2 to 1.2 standard deviations with Whites scoring higher than Blacks. For example, Motowidlo and colleagues reported a Black-White difference of .21 standard deviations in a sample of job incumbents and .38 standard deviations in an applicant sample favoring Whites (Motowidlo et al., 1990). Motowidlo and Tippins (1993) reported a Black-White difference of .38 standard deviation favoring Whites. Pulakos and Schmitt (1996) reported a Black-White difference of .41 standard deviations. These studies used "What would you Most/Least likely do?" response instruction. Weekley and Jones (1999) reported two Black-White difference effect sizes, one of .85 and one of .52 standard deviations using the "Pick the Best and Worst option" response instruction. Using two different selection methods conveying the same test content, Chan and Schmitt (1997) found that the video-based situational judgment test had less adverse impact than the paper-and-pencil test (corrected d = -.28 versus d = -1.19).

From the review of the literature in reducing adverse impact in general and in SJTs in particular, the question of to what extent different response instructions might lead to different subgroup differences has not yet been examined. A recent review of the SJT literature showed a variety of ways to instruct test takers to respond to a situational test item:

- What would you most likely do?
- What would you least likely do?
- What would you most likely do? What would you least likely do?
- Pick the best answer;
- Pick the best answer and then pick the worst answer; and
- Rate each response on effectiveness (McDaniel & Nguyen, 2001).

Of the above response instructions, "What would you most/least likely do" and "Pick the best and worst answer" are the most commonly used in SJT literature. McDaniel and Nguyen (2001) argued that the "Most/Least likely" former instructions would induce a response that is indicative of the respondent's behavioral tendency whereas the "Best/Worst" instructions would induce a response that is indicative of the respondent's knowledge regarding what is the appropriate course of action per given situation. Ployhart and Ehrhart (2001) examined this distinction in response instructions in a study and showed that the behavioral tendency response instruction produced higher validity than that of the knowledge response instruction. However, to what extent the varying response instructions might differentially induce responses indicative of job knowledge has yet to be tested. Also, to what extent different response instructions might lead to differences in adverse impact remains unknown.

Current Study and Hypotheses

In this study, we examined whether different response instructions would influence racial differences in situational judgment test performance. Two response instructions were used, a knowledge instruction (Pick the *Best* and *Worst* option)" and a behavioral tendency instruction ("What would you *Most likely* and *Least likely* do?) We used the exact same test items (i.e., item stem and responses) for both response instructions. This enabled us to control for test content and tease out the effect for response instruction.

We offer several hypotheses. First, personality traits such as conscientiousness and emotional stability have been shown to be valid predictors of job performance across job domains (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Judge, Higgins, Thoresen, & Barrick, 1999). Further, they were shown to have little or no adverse impact as compared to cognitive abilities (Bobko et al., 1999). Whereas both the behavioral tendency SJT and the personality measures are assessments of behavioral tendencies, one would expect the behavioral consistency instruction to have smaller subgroup differences than the knowledge response instruction.

Hypothesis 1: The behavioral tendency response instructions will be associated with smaller subgroup differences than the knowledge response instructions.

Second, there is preliminary evidence that the magnitude of reading comprehension demands inherent in paper-and-pencil situational judgment tests is positively related to the magnitude of racial differences in SJT performance (Chan & Schmitt, 1997; Sacco, Scheu, Ryan, Schmitt, Schmidt, & Rogg, 2000). Given the fact that reading comprehension is a sub-

component of cognitive ability, it is expected that cognitive saturation will predict subgroup performance on situational judgment tests. Therefore, we offered the following hypothesis:

Hypothesis 2: High cognitive saturation of the test will be associated with greater racial differences in situational judgment test performance regardless of instruction set.

Third, job knowledge increases more among those with higher cognitive ability (Schmidt, Hunter, & Outerbridge, 1986) than those with less cognitive ability. Thus, responses from the knowledge instruction should correlate more with cognitive ability and job knowledge than the behavioral tendency instructions. Based on this discussion, we hypothesized:

Hypothesis 3: SJT with a knowledge instruction set (i.e., Best/Worst response instruction) should be more cognitively saturated than SJT with a behavioral tendency instruction set (i.e., Most/Least likely response instruction).

Method

Research Design

A 2x2x2 factorial design was used. The first factor is response instruction: knowledge versus behavioral tendency. We labeled this factor as format factor. The second factor is cognitive loading of the test: high versus low. The third factor is race: White versus Black. The first two factors are within-subject factors and the third factor is the between-subject factor. Hypothesis 1 refers to the possible interaction effect between format and race on SJT test scores. Hypothesis 2 refers to the possible interaction effect between cognitive loading of the SJT and race on SJT test scores. Hypothesis 3 refers to the possible interaction between cognitive loading and the response format of the SJT.

Setting and Participants

Data collected for this study were part of a larger project on a situational judgment test performance. A hundred and sixty-two undergraduate and graduate students from two southern public universities participated in the study. Of these participants, 113 were White and 49 were Black. Participants had a wide age range with a mean of 25.2 (SD = 6.76).

Procedure

A battery of selection tests was administered to groups of participants ranging in size from 3 to 25. Participants signed a consent form and were assigned an identification number at the beginning of the experiment. Participants completed a test battery including a cognitive test (Wonderlic Personnel Test), a Big 5 personality instrument, and two forms of a situational judgment test. Responses were collected anonymously and respondents were asked to respond honestly. Participants were asked to complete the Wonderlic first, followed by the SJT with knowledge instructions, Goldberg's Big 5 personality instrument, and the SJT with behavioral consistency instructions. The Big 5 instrument was administered between the two forms of SJT

to reduce the potential order effect of test session. In hindsight, we realized that we should have counterbalanced the order of test administration (i.e., half knowledge instruction first, half behavioral tendency instruction first). However, previous research showed most test session order effects were small and insignificant (e.g., McFarland & Ryan, 2000). After completing the test battery, participants completed a background measure survey including demographic questions. At the end of the study, participants were debriefed and thanked for their participation. The total testing time was approximately 1 hour.

Measures

Cognitive Ability

Cognitive ability was measured by the Wonderlic Personnel Test (Form A). The test has been used in previous research to measure cognitive ability of adults with test-retest reliabilities above .90 (Wonderlic Inc., 1999).

Work Judgment Survey

The Work Judgment Survey is a situational judgment test described by Smith and McDaniel (1998). The test consists of 31 scenarios or situations, which tap multiple constructs, i.e., conscientiousness, emotional stability, agreeableness, cognitive ability, age, and job experience (Smith & McDaniel, 1998). Due to concern for the length of participation time, four unkeyed scenarios were dropped. For each situation, respondents were asked to select from the five given courses of action. In the knowledge instruction condition, the respondent was asked to identify the best and worst action. In the behavioral tendency condition, the respondent was asked to indicate the responses they would most likely and least likely perform. Smith and McDaniel (1998) used empirical keying in scoring the Work Judgment Survey and we used their key. Specifically, keyed responses were weighted –1, 1, or 2. The endorsement of a non-keyed response received a score of zero. The scale score was the summed scores across items.

The Work Judgment Survey was split into two halves according to the item correlations with the Wonderlic Personnel Test score to constitute the low cognitive loaded and high cognitive loaded sub-tests. Specifically, twelve situations for which the correlation with cognitive ability was above .12 (below which the correlation became non-significant at $p < .05$) formed the high cognitive loaded version for the behavioral tendency instruction. The remaining 15 situations correlating below .12 with cognitive ability formed the low cognitive loaded version of the Work Judgment Survey for the behavioral tendency instruction. For the knowledge instruction condition, 16 items correlating at or above .15 (below which the correlation became non-significant at $p < .05$) with the Wonderlic scores formed the high cognitive (g) loading and 11 items correlating below .15 with the Wonderlic scores formed the low cognitive (g) loading version. The correlation between the low cognitive loaded version of the Work Judgment Survey with the Wonderlic test scores was .14 and .11 for the knowledge and behavioral tendency instructions respectively. The correlation between the high cognitive loaded version of the Work Judgment Survey and the Wonderlic test scores was .42 and .28 for the knowledge and behavioral tendency instructions respectively (See Table 1).

Analyses

We used a doubly repeated 2x2x2 MANOVA procedure to evaluate the effect of race, response instruction and cognitive loading on SJT performance. Response instruction (knowledge vs. behavioral tendency) and cognitive loading (low vs. high) were two within-subjects factors. Race (Black vs. White) was the between-subjects factor. Effect size estimates (*d*) for subgroup differences in situational judgment test performance were computed by subtracting the test mean of White participants from that of Black participants and dividing the difference by the pooled standard deviation. Thus, negative effect sizes indicate that Blacks scored lower than Whites, whereas the positive effect sizes indicate Whites scored lower than Blacks. Sex and race were dummy coded (female = 1, male = 0; White = 1, Black = 2). Effect sizes were corrected for measurement error (i.e., unreliability). We used Cronbach's alpha as an estimate for reliability of SJT in this study.

Missing data

Missing data ranged from 3 to 9 cases for most variables in this study. A linear trend or regression approach was used to impute missing values via SPSS version 9.0.

Table 1. Descriptive statistics and intercorrelations of study variables (N = 162)

	Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1	Gender	.61	.49	-									
2	Age	25.20	6.76	-.13	-								
3	Race	1.30	.46	.17	-.14	-							
4	Cognitive ability	25.70	6.2	-.23	.37	-.49	-						
5	SJT_Knowledge	20.20	5.39	.12	.26	-.21	.39	.66					
6	SJT_Behavioral consistency	14.30	6.54	.02	.23	-.15	.23	.41	.73				
7	High g (Knowledge)	14.20	4.34	.10	.28	-.21	.42	.95	.40	.55			
8	Low g (Knowledge)	6.07	1.91	.10	.09	-.10	.14	.68	.25	.40	.43		
9	High g (Behavioral consistency)	8.02	4.00	.04	.28	-.16	.28	.45	.92	.44	.27	.69	
10	Low g (Behavioral Consistency)	5.92	3.06	.00	.10	-.10	.11	.27	.86	.25	.19	.61	.45

Note. Gender and Race are dummy coded (women = 1, men = 0; White = 1, Black = 2). Cronbach's alpha estimates of reliabilities are underlined; Cognitive ability = Wonderlic test score of participants; SJT Knowledge = Situational judgment test (Best/Worst response instruction); SJT Behavioral Consistency = Situational Judgment test (Most/Least likely response instruction); High g (Knowledge) = High g loading (Best/Worst Instruction Set); High g (Behavioral consistency) = High g loading (Most/Least likely response instruction).

Results

Descriptive statistics and intercorrelations of all variables are shown in Table 1. The MANOVA results are presented in Table 2. Standardized subgroup differences in SJT test performance are shown in Table 3.

As shown in Table 2, the *F* values for all three main effects of race, response instruction, and cognitive loading are significant as well as the *F* values for two of the three two-way interaction terms. The three-way interaction term is not significant. The largest effect was on the cognitive loading factor (MANOVA $F = 404.21$ (1, 160), $p < .01$, Wilk's $\lambda = .284$), indicating that participants' test performance differed greatly in low versus high cognitively loaded SJT. Response instruction factor was also significant (MANOVA $F = 128.017$ (1, 160), $p < .01$, Wilk's $\lambda = .556$). This finding suggests that there is a significant mean difference in the SJT score for knowledge and behavioral tendency instructions. The race factor was significant (MANOVA $F = 7.09$ (1, 160), $p < .05$; Wilk's $\lambda = .958$). This finding suggests that subgroup differences in SJT performance exist in this data set. Table 3 provides further confirmation for this finding. Black-White differences exist in both response instruction sets of the SJT examined in this study. The Black-White difference was ($d = -.45$, $t = 2.57$, $p < .01$) for the knowledge response instruction and ($d = -.32$, $t = 1.96$, $p < .05$) for the behavioral tendency instruction favoring White participants.

There was a significant interaction effect between race and cognitive loading of the test (MANOVA $F = 5.28$ (1, 160), $p < .05$, Wilk's $\lambda = .968$), suggesting that subgroup differences in SJT performance vary as a function of the cognitive loading of the test. The subgroup differences shown in Table 3 further confirm this finding. For the knowledge instruction, the Black-White difference was larger ($d = -.46$) for high cognitive loading and smaller ($d = -.22$) for low cognitive loading. For the behavioral tendency instruction, the Black-White difference was larger ($d = -.34$) for high cognitive loading and smaller ($d = -.22$) for low cognitive loading. Hypothesis 2 was thus supported. Low cognitive saturation of the test was associated with smaller subgroup differences.

Table 2. The effect of Race, Response Instruction, and Cognitive loading on SJT performance

	Main Effects			Interaction Effects			
	Race (1)	Response instruction (2)	Cognitive loading (3)	1x2	1x3	2x3	1x2x3
Situational Judgment	7.09*	128.02**	404.21**	.111	5.28*	240.83**	1.37

Note: *: $p < .05$, **: $p < .01$, two-tailed. The entries are multivariate ANOVA *F*s for SJT performance.

There was a significant interaction effect between response format and cognitive loading of the test (MANOVA $F = 240.83$ (1, 160), $p < .01$, Wilk's $\lambda = .40$), indicating that the pattern of the test associated with different response formats differed greatly when the test is high versus low cognitively saturated. Table 1 further reveals that the correlation between cognitive ability and knowledge SJT scores greater than that and the behavioral tendency SJT scores ($r = .39$ versus $r = .23$, $t = 2.01$, $df = 159$, $p < .05$). Hypothesis 3 was thus supported. SJT with a knowledge instruction set was more cognitively saturated that SJT with a behavioral tendency instruction set.

Contrary to what is expected, the interaction effect between race and response format was not significant, (MANOVA F = .111 (1, 160), $p = .739$, Wilk's $\lambda = .999$), suggesting that subgroup differences do not vary as a function of response format. Table 3 further confirms this finding. Although the Black-White difference in knowledge SJT was larger than that in behavioral tendency SJT, the difference failed to reach statistical significance ($d = -.44$ versus $d = -.34$, corrected for unreliability to be $d = -.55$ versus $d = -.37$, $t = -.77$, $df = 159$, $p = .22$, one-tailed). Thus, Hypothesis 1 was not supported.

Discussion

Despite the fact that SJTs have been increasingly used in practice as a tool to screen job applicants, relatively little is known concerning how to reduce adverse impact inherent in such tests. Chan and Schmitt (1997) showed that the video-based situational judgment test produced less Black-White difference than the traditional paper-and-pencil one. However, the cost involved in producing and administering a video-based method will often make this a less viable option than the traditional paper-and-pencil method. The purpose of this study was to explore the role of response instructions in traditional paper-and-pencil situational judgment test performance: how different response instructions for the exact same test items might lead to different racial differences. This study showed that subgroup differences exist in a situational

Table 3. Effect Size of Subgroup differences in Situational Judgment test performance

Measure	Mean		Standard deviation		Effect size (d)
	Whites (N=113)	Blacks (N=49)	Whites (N=113)	Blacks (N=49)	Blacks vs. Whites
Situational Judgment					
Behavioral tendency instruction					
Low *g* loading	14.96	12.90	6.74	5.86	-.32* (-.37)
High *g* loading	6.12	5.45	3.25	2.54	-.22 (-.33)
	8.43	7.08	3.98	3.93	-.34* (-.41)
Knowledge instruction	20.96	18.57	5.16	5.57	-.45* (-.55)
Low *g* loading	6.19	5.78	1.82	2.08	-.22 (-.34)
High *g* loading	14.77	12.80	4.19	4.41	-.46* (-.62)

Note. Shown in parentheses is Black-White differences corrected for unreliability. *: $p < .05$, two-tailed

judgment test across both knowledge and behavioral tendency response instructions with the behavioral tendency instruction producing somewhat less subgroup difference. The lack of statistical significance might probably have been due to insufficient power due to the small sample size.

The finding that the knowledge response instruction had greater cognitive saturation than the behavioral tendency instruction was important for two reasons. First, because the same test items were used, we could claim that the difference in cognitive saturation was due to the effect of response instruction alone. It strengthened the reasoning that test questions that reflect behavioral tendency should evidence less race-based adverse impact. Second, this finding added further support to the use of this response instruction in practice. Coupled with the finding of higher criterion-related validity (Ployhart & Ehrhart, 2001) associated with the behavioral tendency instruction, practitioners might be advised to use this type of response instruction in applicant screening.

The finding that higher cognitive saturation of the test related to greater subgroup differences deserves special attention from researchers in SJTs for two reasons. First, considering the low to moderate cognitive saturation of the Work Judgment Survey used in this study (r = .39 for knowledge SJT version and r = .23 for the behavioral tendency SJT), a greater racial difference should be expected had the test showed higher cognitive saturation. Second, as McDaniel and Nguyen (2001) noted, there is currently no known technology to build a SJT that controls for cognitive saturation. Stevens and Campion, (1999) attempted to build a SJT having a low correlation with cognitive ability only to find that their tests correlated .81 (uncorrected for attenuation) with cognitive ability. Given this current state of affairs, the behavioral tendency response instruction appeared to have less adverse impact than the knowledge response instruction.

Conclusion

This study contributes to the adverse impact literature in general and SJT literature in particular. As we discussed earlier, previous research reported a wide range of subgroup differences in SJT performance. We showed that response instructions in a SJT might be responsible for that wide range. When respondents were asked to indicate their knowledge, the SJT was more cognitively saturated and displayed greater subgroup differences than when respondents were asked to provide their behavioral tendencies. We hope this study will stimulate future research to address the effects of instruction differences on adverse impact as well as criterion-related and construct validity differences of SJTs.

References

Barrick, M. R., & Mount, M. K. (1991). The Big 5 personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1-26.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52,* 561-589.

Chan, D., & Schmitt, N. (1997). Viedo-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.

Hunter, J. E., & Hunter, R. F. (1984). The validity and utility of alternative predictors of job performance. *Journal of Applied Psychology, 96,* 72-98.

Hurtz, G. M; & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869-879.

Judge, T. A., Higgins, C. A., Thoresen, C. J., Barrick, M. R. (1999). The Big 5 personaltiy traits, general mental ability, and career success across the life span. *Personnel Psychology, 52,* 621-652.

McDaniel, M. A., Morgeson, F. P., Finnegan E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using Situational Judgment Tests: A clarification of the literature. *Journal of Applied Psychology. 86,* 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational Judgment Tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9,* 103-113.

McFarland, L. A., & Ryan, A.M. (2000). Variance in faking across non-cognitive measures. *Journal of Applied Psychology, 85,* 812-821.

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.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66,* 337-344.

Ployhart, R. E., & Ehrhart, M. G. (2001). Effects of response instructions on the criterion-related validity, construct validity, and reliability of situational judgment tests. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion related validity. *Human Performance, 9,* 241-258.

Roth, P. L., BeVier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54,* 297-330.

Sacco, J. M., Scheu, C. R., Ryan, A. M., Schmitt, N., Schmidt, D. B., & Rogg, K. L. (2000). *Reading level and verbal test scores as predictors of subgroup differences and validities of Situational Judgment Tests*. Paper presented at the 15th Annual Meeting of the Society for Industrial and Organizational Psychology. New Orleans, LA.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

Schmidt, F.L., Hunter, J.E. & Outerbridge, A.N. (1986). Impact of job experience and ability on job knowledge, work sample performance and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.

Smith, K. C., & McDaniel, M. A. (April, 1998). *Criterion and construct validity evidence for a situational judgment measure.* Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

SPSS 9.0 for Windows. *Statistical Program for Social Sciences*. SPSS Inc., Chicago, IL

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 25*, 207-228.

Weekley, J.A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.

Weekley, J.A. & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52,* 679-700.

Wonderlic, Inc. (1999). *Wonderlic Personnel Test Manual and Scoring Guide*. Chicago: IL: Author.

Appendix 1

Sample Items from the Work Judgment Survey (Smith & McDaniel, 1998)

1. You are overqualified for your job and this makes you bored and unhappy. In the near future you intend to look for another job but that does not help you right now.
   a. Keep quiet and do your work
   b. Ask your boss for more challenging work
   c. Begin looking around for other jobs for the next year
   d. Involve yourself in enjoyable activities outside work
   e. Just give up and quit.

2. You are in the middle of a difficult job and you ask your boss for help. Your boss won't help.
   a. Get help from someone else
   b. Tell the boss you don't like the boss' attitude
   c. Go to the boss' supervisor and complain
   d. Refuse to do the work
   e. Ask for a meeting with your boss' supervisor

3. You need new equipment and supplies to get the job done right, but your boss does not want to spend the money. The work and morale of your work group are suffering.
   a. Explain the situation to your boss' supervisor
   b. Do your job and mind your own business
   c. Get together with your co-workers and meet with the boss to demand changes
   d. Show the boss how spending money will actually help save money by buying faster equipment, etc.
   e. Spend some of your own money to buy supplies

4. Your boss has demanded you make many changes at once. These changes do not improve performance and everyone is unhappy.
   a. Wait to see what happens.
   b. Get together with some other unhappy employees and complain to the boss
   c. Write up a different plan and present it to the boss
   d. Give the changes time to work and keep a good attitude
   e. Keep doing things the old way

5. Your company has laid off workers. Now you have more work to do.
   a. Work the longer hours it takes to get the job done.
   b. Do the same amount of work you did before
   c. Organize company picnics and social events to improve morale
   d. Look for another job
   e. Try to work harder and smarter by finding faster and simpler ways to get your job done

Author's note: An earlier version of this paper won the Best Paper Award at the Southern Management Association Annual Conference in Atlanta, GA: November 2002.

Author Contact Information

Nhung T. Nguyen
Lamar University
Department of Management & Marketing
4400 Martin Luther King Parkway
Beaumont, TX 77710
Tel.: (409) 880-8295
Fax: (409) 880-8620
E-mail: nguyennt@hal.lamar.edu

Michael A. McDaniel
Virginia Commonwealth University
Department of Management
1015 Floyd Ave.,
Richmond, VA 23284
Tel.: (804) 827-0209
Fax: (804) 828-1602
E-mail: mamcdani@vcu.edu