Applied H.R.M. Research, 2003, Volume 8, Number 1, pages 1-16

A Written Structured Interview by any Other Name is Still a Selection Instrument

Deborah L. Whetzel, Laura E. Baranowski, Jillian M. Petro,
Patrick J. Curtin, and Jennifer L. Fisher
Caliber Associates

The Written Structured Interview (WSI) is a flexible and cost effective alternative to the oral structured interview for collecting detailed information from job candidates. The WSI provides a set of behavioral and/or situational questions to candidates and has them respond in writing in a group setting. The process for developing the instruments for use by a state government, including questions and response rating scales, is described. Data were collected from job incumbents in several state government agencies on the clarity, knowledge, skill, ability and other characteristics (KSAO)-relevance, and job-relatedness of the question, as well as on the effectiveness of behavioral examples of responses. Inter-rater reliability of pilot test responses show an average estimate of .73, indicating that there was a reasonably high level of agreement in the scoring of the WSI responses across a broad range of jobs.

When employers need to hire people who are qualified to perform a job, they frequently screen candidates using a traditional oral face-to-face interview. In fact, the employment interview is one of the most frequently used methods for selecting employees. For example, the Federal Government is one of the largest employers in the United States, currently employing approximately 1.6 million people in over 800 occupations. A survey conducted by the US Merit Systems Protection Board asking Federal managers about the information they consider when assessing candidates for professional and administrative positions revealed that the interview is a "near-universal source of candidate information" (US Merit Systems Protection Board, 2003, p.6). Unfortunately, the oral employment interview process is an expensive, time-consuming way to collect rich, detailed information from candidates. This paper explores the potential of an alternative method for collecting such information, the written structured interview (WSI).

Research has repeatedly shown that when constructed and used appropriately, oral employment interviews are valid for predicting job performance (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Taylor & Small, 2002). However, they can be costly to administer as they require a trained interviewer to conduct and rate each interview. The WSI, rather than requiring face-to-face interviews conducted by trained interviewers, provides a set of questions to candidates and has them respond in writing in a group setting. The WSI is scored later by raters who compare the effectiveness of candidates' responses to behavioral statements describing low, moderate, and high levels of effectiveness of response for each question. This paper describes the development of a WSI that was

accomplished as part of a large-scale effort by a state government to explore the use of new and more efficient selection procedures.

Background

Employment interviews can be structured or unstructured. Structured interviews have a fixed set of questions and each candidate's responses to the questions typically are scored against behaviorally-based rating scales. Unstructured interviews do not use a standard procedure for the administration of questions or scoring. Questions may even vary from candidate to candidate (Huffcutt, Roth, & McDaniel, 1996; Schmidt & Hunter, 1998).

Huffcutt (1992) described four progressively higher levels of interview structure. Level 1 is the typical unstructured interview with no constraints on the questions asked and an evaluation of responses is done without any benchmarks. Level 2 structure imposes limited constraints by specifying the topics to be covered by the questions and establishing some degree of structure on response evaluation. Level 3 requires the prespecification of questions and, although candidates may not be asked precisely the same questions, a defined rating method is used to evaluate responses. Level 4 involves asking candidates precisely the same questions with no deviation and responses are scored according to benchmark answers (e.g., Latham & Saari, 1984; Latham, Saari, Pursell & Campion, 1980). WSIs reflect the Level 4 structure because identical questions are asked of all candidates and the answers are scored according to specified benchmarks.

The content of interview questions similarly falls into two types: behavioral and situational. Behavioral questions (Janz, 1982; Janz, Hellervik, & Gilmore, 1986) ask candidates to think of a time when they had to deal with a particular set of circumstances similar to what they will encounter on the job for which they are being considered, and then to explain those circumstances, their actions, and the resulting outcome(s). These questions take the form of "Tell me about a time when you.…" Behavioral questions are based on the tenet that past job performance predicts future job performance (McDaniel, Schmidt, & Hunter, 1988a; Quinones, Ford & Teachout, 1995). Situational questions present a job scenario that may be encountered by job incumbents and ask candidates to describe how they would deal with the problem or circumstance. These questions take the form of "What would you do if.…?" (Latham & Saari, 1984; Latham et al., 1980). Situational questions are based on the underlying assumption that intentions are related to specific behaviors (Locke, Shaw, Saari, & Latham, 1981). A recent meta-analysis shows that both behavioral and situational questions are valid (Taylor & Small, 2002).

Evidence of Validity for Measures Similar to the WSI

To date, there is no research literature or empirical evidence of the validity of WSIs. However, there are several selection procedures that are similar to WSIs on which there is extensive validity evidence. Oral structured interviews are similar to WSIs in that they ask a candidate the same kinds of questions and responses to

questions are scored using a behaviorally-based rating scale. WSIs only differ from oral interviews in the mode of response (i.e., oral vs. written). Situational judgment tests (SJTs) are similar to WSIs that use situational questions in that scenarios are provided to the candidate, however, they differ in the sense that examinees choose from predetermined alternatives in the SJT (close-ended) rather than generate original answers as in the WSI (open-ended). Training and experience measures (e.g., accomplishment records) are similar to WSIs that use behavioral questions in that they ask candidates about past performance in writing. Given the similarities between WSIs and oral structured interviews, SJTs and training and experience measures, the meta-analytic validity evidence for these three measurement methods is shown in Table 1.

Table 1: Meta-analytic results of three similar assessment methods

	Corrected Correlations	Number of Subjects in Studies	90% Credibility Values
Oral Structured Interviews[a]	.51	12,847	.07
Situational Judgment Tests[b]	.34	10,640	.16
Training and Experience (Behavioral Consistency)[c]	.45	1,148	.33

The oral structured interview is a selection procedure designed to predict future job performance on the basis of candidates' oral responses to oral inquiries (McDaniel et al., 1994). WSIs can use the same kinds of questions as oral interviews.

An example of a behavioral question is:

> Tell me about a time when you had to develop action plans to accomplish operational objectives within the constraints of available resources.

Related behavioral examples are shown below:

- *Low*: Developed plans based only on instincts and own past experience, rather than on available data; created action plans without soliciting staff or other input

[a] Oral structured interview data are from "The validity of employment interviews: A comprehensive review and meta-analysis," by McDaniel, Whetzel, Schmidt, & Maurer,1994, *Journal of Applied Psychology, 79,* p. 606.

[b] Situational judgment test data are from "Use of situational judgment tests to predict job performance: A clarification of the literature," by McDaniel, Morgeson, Finnegan, Campion, and Braverman, 2001, *Journal of Applied Psychology, 86,* p. 735.

[c] Training and experience data are from "Job experience correlates of job performance," by McDaniel, Schmidt, & Hunter, 1988, *Journal of Applied Psychology, 73,* p. 300.

- *Moderate*: Incorporated the use of various resources and their availability when developing plans; solicited input from staff when creating plans

- *High*: Created contingency plans so that changes can be incorporated into the original plan; incorporated input from staff and outside sources (other affected departments) when creating plans

An example of a situational question is:

> What would you do if you had to train people as soon as possible and you did not have the expertise to provide the training yourself?

Related situational examples are shown below:

- *Low*: Have new employees "learn it on their own" on the job; perform work on own that could be delegated if employees were trained

- *Moderate*: Read a book and learn the subject area; identify and suggest training needs for new employees in various areas of expertise

- *High*: Identify individuals who are knowledgeable in the subject area and ask them to teach the new people; provide developmental opportunities to new employees who wish to learn a new function

As shown in Table 1, McDaniel et al. (1994) found an average corrected validity of .51 using 12,847 subjects across studies for oral structured interviews using criteria collected for research purposes only.

SJTs are similar to WSIs in that they ask candidates to describe how they would respond in a specific situation (Motowidlo, Dunnette & Carter, 1990). Most widely used SJTs have a multiple-choice format and candidates are asked to identify the most/least effective response option for a given situation. McDaniel, Morgeson, Finnegan, Campion, & Braverman (2001) found an average corrected validity of .34 using 10,640 across studies for SJTs that they described as useful and generalizable.

Training and experience measures are similar to WSIs because they require candidates to describe their major achievements in several job-related areas. For example, the accomplishment record (Hough, 1984; Hough, Keyes, & Dunnette, 1983), a type of training and experience measure, asks job candidates to provide written descriptions of their accomplishments for several knowledge, skills, abilities and other characteristics (KSAOs) or performance dimensions. Each accomplishment contains a description of the situation, the candidate's actions, and the outcome of the candidate's actions. Trained raters evaluate the written responses against a scoring rubric. For the WSI, candidates are asked to think of a specific

time when they had to use a job-relevant KSAO. Then, they are asked to describe the situation, the actions they took, and the outcome of their actions. Research suggests that accomplishment records can be scored reliably and that their scores have been useful predictors of job performance. McDaniel, Schmidt, & Hunter (1988b) found a corrected validity of .45 using 1,148 subjects across studies for behavioral consistency methods of training and experience measures that are similar to the accomplishment record.

Boundary Conditions for Structured Interview Validity

Schmidt and Rader (1999) explored the boundary conditions for interview validity using telephone interviews. Despite the differences between face-to-face interviews and telephone interviews, Schmidt and Rader found that the telephone interview had essentially the same level of criterion-related validity for predicting job performance as the traditional face-to-face interview. They explain their results by suggesting that different types of structured interviews measure constructs with known generalizable validity (e.g., conscientiousness and general mental ability). Therefore, to the extent that similar constructs are measured, similar validities should result. An important advantage of phone interviews over face-to-face interviews is that the effect of interviewer bias based on factors such as appearance and background is reduced. This advantage also applies to the WSI. In sum, the telephone interview was found to be nearly as valid as the oral face-to-face interview and the telephone interview shares several advantages with the WSI over the oral face-to-face interview. Thus, written responses may be just as valid for predicting performance as oral responses.

Motowidlo, Carter, Dunnette, Tippins, Werner, Burnett and Vaughn (1992) also examined conditions that might affect interview validity. They conducted a series of studies investigating structured behavioral interviews. In one of their studies, they had students write summaries of interview responses and had readers rate the written responses on a set of dimensions. Their results showed that nonverbal cues, present for oral interviews, were not necessary for the validity of the structured behavioral interview and that valid interview judgments can be made from the content of the interview answers alone.

Method

As part of a test development project for a state government, we examined the potential of a WSI for selecting candidates. As a first step in the process, we analyzed six state government jobs to identify knowledge, skills, abilities and other characteristics (KSAOs) that might be appropriately measured using a WSI. The jobs were Accountant, Communications Technician, Electronics Technician, Geologist, Graphic Arts Operator, and Public Information Specialist. Six KSAOs were found to be common across many of these jobs. They were:

- Agreeableness
- Conscientiousness

- Stress tolerance
- Ability to generate new ideas and organize and analyze them logically
- Ability to manage own time and coordinate with others
- Ability to work without close supervision.

Initially, we wrote four or five behavioral or situational interview questions for each KSAO resulting in a total of 29 questions. We developed both kinds of questions because behavioral questions allow candidates to reflect on past behavior (past behavior is a predictor of future behavior) and situational questions allow candidates to project their behavior to situations likely to occur on the job (intentional behavior is highly correlated with actual behavior). We also developed 6-26 behavioral examples of responses that we believed described low, moderate, or high levels of effectiveness for each question.

Ratings of Questions

In a series of workshops, 61 subject matter experts (SMEs) from several state agencies rated the 29 questions for clarity, job-relevance and KSAO-relatedness. SMEs were asked "Is the question clear?" and they provided a rating of "0" for No or "1" for Yes. SMEs also were asked "Is the question job-relevant?" and they provided ratings on a five-point scale, with a "5" being the most job-relevant. SMEs then were asked "Is the question KSAO-related?" and they provided ratings on a five-point scale, with a "5" being the most KSAO-related.

Ratings of Behavioral Examples

To obtain response effectiveness ratings, we created a rating scale that would provide behavioral examples at each level of effectiveness for each question. The scale consisted of behaviors that define low, moderate and high levels of responses. To identify which behaviors belonged at each level, we created a survey in which the behavioral examples were randomly ordered for each question. SMEs rated each behavioral example on a 5-point scale, with a "5" being the most effective.

To determine the placement of each example (i.e., low, moderate, or high) for each question, we analyzed the effectiveness ratings given by the SMEs. We selected examples for each scale by using only those examples that had relatively high agreement (i.e., a standard deviation on effectiveness of 1.25 or less). We analyzed the responses for each question separately such that the lowest rated statements represented "Low", the next lowest statements represented "Moderate", and the highest rated statements represented "High". In this way, all "Low" examples were lower than "Moderate" examples and "Moderate" examples were lower than "High" examples for each question. The final scales had at least three behavioral examples for each level of performance for each question.

Reliability of Scales

To determine the reliability of the scales, we selected one question for each KSAO (for a total of six questions) and asked 6-14 SMEs in each of three jobs to write an answer to each question. The SMEs participating in this exercise were Public Information Specialists, Geologists, and Accountants. They generated 64 responses to the six questions.

Five psychologists independently rated the 64 responses to the WSI questions using the rating scales. We used these ratings to compute reliability as described below.

Results

Ratings of the Questions

The results of SME ratings on clarity of the question are shown in Table 2. For 25 of the 29 questions, mean ratings of 1.0 were obtained for most jobs, indicating widespread agreement that the questions were clear.

Table 3 shows the mean ratings for job relevance of the questions. For 22 of the 29 questions, average job relevance ratings were above 3.0 across all six jobs, indicating that the questions were viewed by SMEs as being relevant across jobs.

Table 2: Mean Ratings for Question Clarity

KSAO: Question	*Job A*[a]	*Job B*[b]	*Job C*[c]	*Job D*[d]	*Job E*[e]	*Job F*[f]
Agreeable						
Question 1	0.69	0.80	0.86	0.50	1.00	0.63
Question 2	1.00	0.83	1.00	1.00	1.00	1.00
Question 3	1.00	1.00	1.00	1.00	1.00	1.00
Question 4	1.00	1.00	1.00	1.00	1.00	0.63
Question 5	1.00	0.80	1.00	1.00	1.00	0.88
Conscientious						

[a] Job A is the accountant job family, consisting of Staff Accountants (N=7); Senior Accountants (N=6); Accounting Managers (N=6); Audit Managers (N=5); and Accounting Director IIIs (N=6).

[b] Job B is the communications technician job family consisting of Communications Technicians IIs (N=4) and Communications Technician Supervisors (N=2).

[c] Job C is the electronics technician job family consisting of Electronics Technicians (N=4), Electronics Shop Supervisor (N=1) and supervisors of Electronics Shop Supervisors (N=2).

[d] Job D is the geologist job family, consisting of Geologist Is (N=2), and Geologist IIs (N=3).

[e] Job E is the Graphic Arts Operator Supervisor job (N=5).

[f] Job F is the Public Information Specialist job (N=11).

Table 2: Mean Ratings for Question Clarity

KSAO: Question	*Job A*[a]	*Job B*[b]	*Job C*[c]	*Job D*[d]	*Job E*[e]	*Job F*[f]
Question 1	1.00	--	0.71	1.00	1.00	1.00
Question 2	1.00	--	1.00	1.00	1.00	0.82
Question 3	1.00	--	1.00	1.00	1.00	0.91
Question 4	0.95	--	1.00	1.00	1.00	1.00
Question 5	1.00	--	1.00	1.00	1.00	0.91
Generate new ideas						
Question 1	0.95	1.00	1.00	1.00	--	1.00
Question 2	0.95	1.00	1.00	1.00	--	1.00
Question 3	0.89	1.00	1.00	1.00	--	0.90
Question 4	1.00	1.00	0.75	1.00	--	0.73
Question 5	1.00	1.00	0.75	1.00	--	0.82
Manage own time						
Question 1	1.00	1.00	0.86	1.00	1.00	1.00
Question 2	0.95	1.00	1.00	1.00	1.00	1.00
Question 3	1.00	1.00	0.71	0.80	1.00	1.00
Question 4	0.86	1.00	0.57	1.00	1.00	0.91
Question 5	0.95	1.00	1.00	0.80	1.00	0.91
Tolerate stress						
Question 1	1.00	1.00	1.00	--	1.00	1.00
Question 2	0.95	1.00	0.75	--	1.00	1.00
Question 3	1.00	1.00	1.00	--	1.00	0.75
Question 4	0.86	1.00	0.75	--	1.00	0.88
Question 5	1.00	1.00	1.00	--	1.00	0.86
Work without close supervision						
Question 1	1.00	0.75	1.00	1.00	--	1.00
Question 2	1.00	1.00	0.86	1.00	--	1.00
Question 3	1.00	1.00	0.86	1.00	--	0.80
Question 4	1.00	1.00	1.00	1.00	--	1.00

Table 3: Mean Ratings for Job Relevance

KSAO: Question	*Job A*[a]	*Job B*[b]	*Job C*[c]	*Job D*[d]	*Job E*[e]	*Job F*[f]
Agreeable						
Question 1	3.92	3.17	3.50	1.00	4.20	3.25
Question 2	4.46	3.33	4.14	3.00	4.40	3.50
Question 3	3.46	3.33	3.43	3.00	3.80	3.43
Question 4	3.46	2.83	3.57	4.00	4.60	3.38
Question 5	3.85	3.17	3.57	4.00	3.60	3.25
Conscientious						
Question 1	3.57	--	2.86	3.67	1.67	3.36
Question 2	4.24	--	3.86	3.00	3.67	3.55
Question 3	4.43	--	3.29	3.00	3.00	3.91
Question 4	4.48	--	3.57	3.00	3.67	4.00
Question 5	4.65	--	4.00	4.00	3.33	4.45
Generate new ideas						
Question 1	3.58	4.33	3.50	3.40	--	3.64
Question 2	3.79	4.17	3.25	3.20	--	3.45
Question 3	3.95	4.00	3.00	3.40	--	3.55
Question 4	4.05	4.17	4.00	4.20	--	3.82
Question 5	4.11	4.00	3.33	3.60	--	3.45
Manage own time						
Question 1	4.50	3.75	4.43	4.20	4.00	4.64
Question 2	4.36	4.00	4.00	3.80	4.00	4.18
Question 3	4.05	3.50	3.71	3.80	3.33	4.27
Question 4	3.95	3.75	3.50	3.60	3.00	4.00
Question 5	4.23	3.50	3.86	3.60	3.00	4.27
Tolerate Stress						
Question 1	4.14	3.75	3.50	--	2.67	4.13
Question 2	3.95	4.50	3.75	--	3.00	4.63
Question 3	3.76	2.25	3.50	--	2.67	3.88
Question 4	3.30	2.50	2.75	--	2.00	3.38
Question 5	3.38	2.75	3.00	--	2.33	3.38
Work without close supervision						
Question 1	4.59	3.75	3.86	3.50	--	4.20
Question 2	4.09	4.00	4.14	2.50	--	4.40
Question 3	4.09	4.00	3.86	4.00	--	4.40
Question 4	4.29	4.50	3.86	4.00	--	4.00

Table 4 provides the average ratings for KSAO-relatedness. For 21 of the 29 questions, average KSAO-relatedness ratings were above 3.0 across all six jobs, indicating that there was agreement among SMEs that the questions were related to the KSAO for which they were written.

Table 4: Mean Ratings for KSAO -Relatedness

KSAO: Question	*Job A*[a]	*Job B*[b]	*Job C*[c]	*Job D*[d]	*Job E*[e]	*Job F*[f]
Agreeable						
Question 1	3.83	3.00	3.67	3.00	3.20	3.50
Question 2	4.08	2.83	4.57	3.50	3.80	4.00
Question 3	3.54	3.00	4.14	2.00	4.00	3.71
Question 4	3.62	3.17	3.86	2.50	4.00	3.50
Question 5	3.85	3.17	3.71	2.50	3.60	3.88
Conscientious						
Question 1	3.62	--	3.57	4.00	2.33	3.64
Question 2	4.24	--	4.43	4.00	3.33	3.64
Question 3	4.57	--	4.14	4.33	3.33	4.27
Question 4	4.52	--	4.00	4.00	4.00	4.45
Question 5	4.60	--	4.14	4.33	3.33	4.55
Generate new ideas						
Question 1	3.79	4.17	4.00	4.20	--	4.36
Question 2	3.74	3.83	4.00	4.00	--	4.09
Question 3	4.00	4.00	3.75	4.20	--	4.36
Question 4	4.00	4.17	4.25	4.20	--	4.00
Question 5	4.06	4.17	4.00	4.00	--	3.91
Manage own time						
Question 1	4.64	3.75	4.86	4.60	4.00	4.73
Question 2	4.43	4.00	4.29	4.00	4.00	4.36
Question 3	4.14	3.50	4.14	4.40	3.67	4.64
Question 4	4.00	4.00	4.17	4.40	3.00	4.27
Question 5	4.27	3.50	4.14	4.40	3.00	4.45
Tolerate stress						
Question 1	4.14	3.75	3.25	--	3.00	4.38
Question 2	4.00	4.50	4.25	--	3.33	4.71
Question 3	3.80	2.25	4.00	--	3.00	4.00
Question 4	3.45	2.50	3.25	--	2.00	3.75
Question 5	3.48	3.50	4.75	--	2.00	3.75
Work without close supervision						
Question 1	4.59	3.75	4.71	4.50	--	4.80
Question 2	4.14	4.25	4.14	3.50	--	4.40
Question 3	4.18	4.00	4.29	4.50	--	4.60
Question 4	4.33	4.50	4.43	4.50	--	4.60

Ratings of the Behavioral Examples

We calculated inter-rater reliabilities using the Shrout and Fleiss (1979) ICC (3,k) formula, which is equivalent to Cronbach's coefficient alpha (Cronbach, 1951). Table 5 shows for each KSAO, the number of behavioral examples rated, the number of raters, and the inter-rater reliability for each KSAO. Between 20 and 37 raters provided ratings of effectiveness for the various behavioral examples for each question. Using all raters, inter-rater reliabilities ranged from .88 to .99, indicating a high degree of reliability.

Table 5: Inter-rater reliability of effectiveness ratings

KSAO	*Number of Behavioral Statements*	*Number of Raters*	*Inter-rater Reliability*
Agreeable	13-18	20	.96 - .98
Conscientious	6-10	30	.97 - .99
Generate New Ideas	18-26	33	.94 - .97
Manage Own Time	12-20	37	.88 - .98
Tolerate Stress	17-24	32	.99
Work Without Close Supervision	13-20	34	.98 - .99

Reliability of the Scales

We also calculated the reliability of the 64 responses to the six questions using the formula described above. The reliability was based on all five psychologists' ratings. The inter-rater reliability ranged from .62-.90 with a total across items and raters of .87, as shown in Table 6. Because two raters likely will be used to score the WSI when it is implemented by the state government, we adjusted the reliability using the Spearman-Brown formula. Since the "test length" (i.e., number of raters) was reduced from five to two (a factor of .4), the reliability was calculated as .73.

Table 6: Inter-rater reliability of response ratings (5 raters)

KSAO	*Number of Responses Scored*	*Reliability*
Agreeable	8	.89
Conscientious	10	.76
Generate New Ideas	12	.75
Manage Own Time	14	.90
Tolerate Stress	9	.62
Work Without Close Supervision	11	.81
Overall items and raters	64	.87

Discussion

Advantages of the WSI

The WSI provides several practical advantages over the oral structured interview. First, it can be administered to large groups of candidates simultaneously. This can result in cost savings for employers over the traditional face-to-face interview. Applicants in widely scattered geographical locations can be interviewed with no travel costs for interviewers or interviewees. If an employer has offices in several cities, applicants at all locations can be interviewed from one central office.

Second, the consistency of administration across candidates is likely to increase its reliability. Using the exact same questions asked of all examinees reduces the possibility that extra follow-on or prompting questions could affect examinees' responses.

Third, it reduces the effect of interviewer bias on factors such as appearance, accent, and background, which would increase the objectivity of the interview. In addition, the legal defensibility of the process would be enhanced due to the absence of these potential biases.

Fourth, it requires no interviewer, which eliminates the need for interviewer training and skill building. However, this cost would likely be offset by the need for trained scorers.

Disadvantages of the WSI

In spite of the practical advantages of the WSI, there are some caveats to using the WSI. First, WSI responses may elicit bias in rater scoring attributable to candidate handwriting style and quality. This issue could be overcome by using computers for interview administration. Computer administration of WSIs, however, may introduce error associated with differences in candidate exposure to computers and various kinds of software.

Another possible criticism of WSIs is that nonverbal cues are unavailable. On the basis of a study in which readers rated written answers to questions, Motowidlo et al. (1992) concluded that valid judgments about interviewees can be made even when judges do not have access to information about their appearance, behavior, or vocal characteristics as they have in the face-to-face interview. Their study was designed to filter out such nonverbal cues and leave the content of the interviewees' answers as the only remaining source of information in the interview. They concluded that, because the content of answers is determined by the standard questions asked of all interviewees, it is the interview's structured format that is a source of its validity.

Another potential problem with the WSI is that because there is likely to be a higher g-loading on the instrument, because of the writing component, there may be greater adverse impact on minority groups. Huffcutt, Roth, McDaniel (1996) and McDaniel et al. (1994) presented meta-analytic evidence that oral structured interview scores correlate positively with general mental ability. This relationship is likely to be true for the WSI as well. Future research might describe the factor

structure underlying the WSI, as compared to that underlying the oral structured interview, particularly the g-loading of WSIs, given the implications for adverse impact.

Limits to the Generalizability of This Study

There are several limitations of this study that can affect the possible generalization of these results to an operational environment. Concerning the estimate of reliability, the responses used for the reliability analyses were generated by job incumbents, not actual candidates. Job incumbents are more likely to know the "right" answers or at least be more familiar with agency practice. Thus, their responses are likely to be more similar to each other and to be illustrative of behaviors in the High category. Additionally, the effectiveness of their responses were rated by psychologists with training and experience in developing selection methods. To the extent that ratings often are made by personnel analysts in operational settings, these results may overestimate the reliability of the WSI.

Recommendations for Operational Use of the WSI

First, WSIs should be used for jobs that require writing skills. Even if writing is not a dimension measured by the WSI, use of a WSI for a job that does not require writing skills introduces error variance, reduces face validity, and makes the organization more susceptible to legal challenge.

Second, test security concerns need to be considered. Although examinees are more likely to remember questions, parallel forms can be constructed and questions can be pre-tested to determine their difficulty by administering non-scoreable items as part of the scored WSI. Once data are collected on the unscored items, they can be scored operationally. Equally difficult questions measuring the same KSAO could then be administered during a test session or across sessions.

References

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests, *Psychometrika, 16,* 297-334.

Hough, L. M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69*, 135-146.

Hough, L. M., Keyes, M. A., & Dunnette, M. A. (1983). An evaluation of three "alternative" selection procedures. *Personnel Psychology, 36*, 261-275.

Huffcutt, A.I. (1992). *An empirical investigation of the relationship between multidimensional degree of structure and the validity of the employment interview.* Unpublished doctoral dissertation, Texas A&M University, College Station.

Huffcutt, A.I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184-190.

Huffcutt, A. I., Roth, P., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473.

Janz, J.T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577-580.

Janz, J.T., Hellervik, L., & Gilmore, D.C. (1986). *Behavior description interviewing: New, accurate and cost effective*. Newton, MA: Allyn & Bacon.

Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology, 69*, 569-573.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.

Locke, E.A., Shaw, K.M., Saari, L.M., & Latham, G.P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin, 90,* 125-152.

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.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988a). Job experience correlates of job performance. *Journal of Applied Psychology, 73*, 327-330.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988b). A meta-analysis of the validity of methods for training and experience in personnel selection. *Personnel Psychology, 41,* 283-314.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599-616.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77*, 571-587.

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.

Quinones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48*, 887-910.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

Schmidt, F.L. & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology, 52*, 445-464.

Shrout, P. E. & Fleiss, J. L. (1979). Intra-class correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Taylor, P.J. and Small, B. (2002). Asking applicants what they *would* do versus what they *did* do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277-294.

US Merit Systems Protection Board (2003). *The Federal selection interview: Unrealized potential*. Washington, DC: Government Printing Office.

Author Notes

1. The authors would like to thank Jennifer Harvey and Teresa Russell for their work in developing the interview questions, rating scales, and data collection instruments used in this study. We also would like to thank Daniel Felker, Lance Anderson, and Kenneth Graham for their thoughtful comments on early drafts of this manuscript.
2. Portions of this paper were presented at the 18th annual conference of the Society of Industrial and Organizational Psychology, Orlando, FL (April, 2003)
3. Correspondence regarding this article should be sent to Deborah L. Whetzel, Caliber Associates, 10530 Rosehaven Street, Suite 400, Fairfax, VA 22030. Electronic mail may be sent to whetzeld@calib.com. Phone: (703) 219-4402. Fax: (703) 219-3777.