Structured Interviewing: A Note on Incremental Validity and Alternative Question Types

Michael A. Campion, James E. Campion, and J. Peter Hudson, Jr.

The study assesses whether a structured interview can have incremental validity in the prediction of job performance beyond a battery of cognitive ability tests and whether future-oriented (e.g., situational) or past-oriented (e.g., behavior description) questions have higher validity. A 30-item structured interview, with 15 future and 15 past questions, and a battery of nine tests were correlated with job performance in a sample of 70 pulp mill employees. All measures exhibited high variance and high reliability. Uncorrected validities were .50 for the interview and .46 for the tests, and the interview had incremental validity beyond the tests. The validity for past questions (.51) was higher than for future questions (.39), but not significantly higher. However, past questions showed incremental validity beyond future questions (but not vice versa), and both question types had incremental validity beyond the tests.

Historically, the employment interview was viewed as lacking reliability and validity due partly to its lack of standardization (Arvey & Campion, 1982; Harris, 1989; Schmitt, 1976). There has been substantial recent effort to increase standardization by structuring the process. Two meta-analyses have concluded that structured interviews are more valid than unstructured interviews (Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989).

There are many ways interviews can be structured (Campion, Pursell, & Brown, 1988)—questions based on a job analysis, same questions asked of all candidates, anchored rating scales for scoring answers, notetaking, a panel of interviewers, and consistent administration (e.g., control of procedures, prompting, and extraneous influences).

Research has raised many questions about why structured interviews are effective. This study addresses two of these questions. One question is whether structured interviews have validity simply because they assess cognitive ability. Cognitive ability tests may be the most robust predictors of job performance (Hunter & Hunter, 1984; Reilly & Chao, 1982), and they are inexpensive and readily available. Not only is the relationship to such tests important for understanding the construct validity of structured interviews but also, when structured interviews are used in conjunction with tests, it is desirable that they have incremental validity beyond tests to justify the additional effort and expense in their development and administration.

Three previous studies have examined relationships with cognitive ability tests. Campion et al. (1988) found moderate relationships between an interview and four tests (average r = .43). But despite substantial validity for the interview (uncor-

Michael A. Campion, Krannert School of Management, Purdue University; James E. Campion, Department of Psychology, University of Houston; J. Peter Hudson, Jr., Jeanneret & Associates, Inc., Houston, Texas.

Correspondence concerning this article should be addressed to Michael A. Campion, Krannert School of Management, Purdue University, West Lafayette, Indiana 47907-1310.

rected r = .34, corrected r = .56), there was no incremental validity beyond the tests. It was speculated that the interview was only effective because it operated like an "orally administered cognitive ability test" (p. 36).

Walters, Miller, and Ree (1993) found validity for a structured interview (R = .21), but no incremental validity beyond cognitive tests. Correlations with the tests were not reported, but the authors commented that interviews and tests "measure, in some degree, the same construct" (p. 36).

Motowidlo et al. (1992) found only small correlations with tests and other cognitive indicators such as grades and class rank (average r = .15). They concluded that the interviews were "not heavily saturated with constructs such as cognitive ability" (p. 584), but they did not explicitly test whether the validity of their interviews (average r = .22) contributed incrementally beyond the tests. These studies are also hard to compare because they differed on many important factors (e.g., type of questions, degree of structure, setting, jobs, sample heterogeneity, etc.).

The first purpose of this study is to further examine the incremental validity of a structured interview in the prediction of job performance beyond a battery of cognitive ability tests. It differs from previous studies in that the interview is designed to complement rather than duplicate the tests. For example, the interview in Campion et al. (1988) had items on reading, math, and mechanical knowledge, whereas the interview here examines teamwork, self-management, commitment, and other social attributes that are not usually viewed as cognitive ability but that may be areas where an interview might have unique measurement value (Schmitt, 1976). Further, this study uses nine cognitive tests to provide a rigorous assessment of incremental validity.

Another issue is question type, with two types emerging. Situational interviews ask future-oriented questions. They pose hypothetical situations that may occur on the job, and candidates are asked what they would do (Campion et al., 1988; Latham, Saari, Pursell, & Campion, 1980; Pursell, Campion, & Gaylord, 1980; Weekley & Gier, 1987). Situational questions may predict because of the relation between goals or intentions

and future behavior (Locke & Latham, 1984). Behavior description interviews ask past-oriented questions. They ask candidates to describe what they did in past jobs as it relates to requirements of the job they are seeking (Janz, 1982; Motowidlo et al., 1992; Orpen, 1985). Behavior description questions may be predictive because past behavior predicts future behavior.

The second purpose is to examine whether past or future questions have higher validity. This issue has not been addressed in the literature. Previous research cannot address this issue because question type was confounded with degree of structure. Situational interviews have been highly structured, and behavior description interviews have been more flexible. Although both are based on job analyses, all candidates are asked the same questions in situational interviews, whereas interviewers select questions from an array or pattern in behavior description interviews. Although both use anchored rating scales, each answer is rated separately after it is given in situational interviews, whereas all judgments are made at the end based on all answers together in behavior description interviews. Finally, situational interviews use panels of interviewers, unlike behavior description interviews. To allow an unconfounded comparison, this study holds these factors constant by using the more highly structured format for both question types.

Method

Setting and Sample

The setting was a Southeastern pulp mill. The selection system was used to staff a new pulp mill that was being built to replace an existing mill. Candidates were employees in the old mill. Selection procedures included a structured interview, tests, and job performance. The setting allowed a concurrent validation in that the interview and tests could be validated against job performance. This was possible because job requirements in the new mill were similar to those in the old mill, and selection procedures focused on attributes of current jobs that would be needed to a greater degree in the new jobs. The higher level of requirements on the new jobs justified the need for a selection system. There were two advantages to this setting. First, there was high test-taking motivation because scores would determine who received the internal promotion opportunities. Second, there was no restriction of range due to preselection.

The sample consisted of 70 employees, including 31 Blacks and 4 women. Average tenure was 21.7 years ($SD = 8.29$) and average education was 11.9 years ($SD = 1.62$). Statistical power was 82% to detect an observed (uncorrected) correlation of .30 ($p < .05$, 1-tailed).

Job Analysis

Because the new pulp mill did not yet exist when hiring decisions had to be made, the job analysis consisted of projecting changes from current job descriptions, reviewing job descriptions from company locations with similar new pulp mill equipment, examining planning documents, conducting discussions with job experts, and collecting critical incidents. The result was two sets of knowledge, skill, ability, and other requirement (KSAOs) lists, on which selection procedures were based: (a) 10 cognitive abilities that would be assessed by the tests (e.g., reading, writing, arithmetic, problem solving, mechanical knowledge, perceptual abilities, etc.) and (b) 17 attributes not apparently reflecting cognitive abilities that would be assessed by the interview (e.g., initiative, teamwork, resolving conflict, commitment to improvement, work ethic, safety orientation, accepting responsibility, growth orientation,

leadership, etc.). These KSAOs were needed on current jobs as evidenced by the fact that they came from current job descriptions (a above) or from critical incidents with reference to current jobs (b above), but were judged by experts to be needed to a greater degree on the new jobs.

Structured Interview

The approach to developing a highly structured interview described by Campion et al. (1988) was used. Thirty items were developed based on job analysis information, with 15 future- and 15 past-oriented. Five-point rating scales were developed for each question with the 5, 3, and 1 points anchored with definitions, descriptions, and example answers.

To check that future and past questions had comparable content validity, three analysts naive to the purposes of the study linked KSAOs with each question and its scale. Using only linkages with agreement between two or more analysts, future questions linked with 13 and past questions with 12 of the 17 KSAOs. Eleven were with the same KSAOs. The question types also linked to KSAOs in similar proportions. The correlation across the 17 KSAOs between number of links with future questions and number with past questions was .70 ($p < .05$). Future questions showed a larger number of total links with the KSAOs than past questions, however (34 vs. 25).

Future and past questions were of comparable length and administered in the same manner by the same interviewers. Examples of each question type are shown in Table 1.

Seven managers served as interviewers in panels of two (50% of cases) or three. In no case did the panel include a candidate's supervisor. Standardization was enhanced by asking each candidate all 30 questions in the same order, having one interviewer ask all the questions, not allowing prompting and follow-up questions, and having all panel members independently record and rate each answer immediately after it was given. Although interviewers were experienced with structured interviewing, a 1-day refresher training program was conducted that included a review of the process and three practice interviews. Finally, the future questions were asked before the past questions because it was thought to be easier for candidates to begin with hypothetical questions. This was expected not to influence the results, because (a) the interview

Table 1
Examples of Structured Interview Questions

Future-oriented question: Suppose you had an idea for a change in work procedure to enhance quality, but there was a problem in that some members of your work team were against any type of change. What would you do in this situation?

    (5) Excellent answer (top third of candidates)—Explain the change and try to show the benefits. Discuss it openly in a meeting.
    (3) Good answer (middle third)—Ask them why they are against change. Try to convince them.
    (1) Maginal answer (bottom third)—Tell the supervisor.

Past-oriented question: What is the biggest difference of opinion you ever had with a co-worker? How did it get resolved?

    (5) Excellent answer (top third of candidates)—We looked into the situation, found the problem, and resolved the difference. Had an honest conversation with the person.
    (3) Good answer (middle third)—Compromised. Resolved the problem by taking turns, or I explained the problem (my side) carefully.
    (1) Marginal answer (bottom third)—I got mad and told the co-worker off, or we got the supervisor to resolve the problem, or I never have differences with anyone.

Note. Both questions are intended to assess conflict resolution and collaborative problem-solving knowledge, skills, and other requirements.

was so highly structured, (b) the interviewers were trained to establish rapport and put the candidate at ease, and (c) there was no correlation between item order and item validity within each question type. The total score was the average across questions and interviewers. Coefficient alpha reliability was .93, interrater reliability of the mean of the panel was .97, and average absolute interrater differences among interviewers was .16 (on the 1 to 5 scale). Thus, the interview showed high reliability and agreement.

Cognitive Ability Tests

Nine tests were chosen by job experts to represent the types and levels of cognitive abilities identified in the job analysis and traditionally assessed in employment testing. Three assessed verbal ability: Ramsay's (1991) Job Skills Reading Test (Form A) measured reading comprehension; the Personnel Tests for Industry (Verbal, Form A) measured vocabulary (Wesman, 1980); and Flanagan's (1975) Industrial Tests (Expression) measured grammar. Three assessed math ability: Science Research Associates' (1986) Arithmetic Index measured arithmetic; the Psychological Services Inc. Basic Skills Tests (Problem Solving, Form A) measured mathematical problem solving (Ruch, Weiner, McKillip, & Dye, 1985); and Flanagan's (1975) Industrial Tests (Scales) measured the ability to read measurement scales and graphs. One assessed mechanical ability: the Differential Aptitude Test measured mechanical reasoning (Bennett, Seashore, & Wesman, 1991). Two assessed perceptual ability: the Employee Aptitude Survey #4 (Form A) measured visual speed and accuracy and #3 (Form A) measured visual pursuit (Ruch & Ruch, 1980).

Total scores on all tests were the number correct. Coefficient alpha cannot be used to estimate reliability for speeded tests. Where other forms of reliability were reported in the manuals (e.g., alternative forms), the levels were acceptable. To provide an overall index of cognitive ability, a test composite was formed by converting to z scores and then averaging across the nine tests. Coefficient alpha reliability (appropriate when applied to the composite) was .93. The tests were administered by a trained personnel representative.

Job Performance

A 6-item measure was developed to assess job performance in terms of the KSAOs identified in the job analysis. Two items reflected the cognitive areas: learning orientation and technical knowledge. Four items reflected other areas: self-management, team contribution, communication, and quality and continuous improvement. Each item was defined by a brief explanation and coupled with a 5-point rating scale ranging from 5, *well above average (top 20% of employees)* to 1, *well below average (bottom 20% of employees)*.

To enhance reliability and accuracy, two (59% of cases) or three (27%) independent evaluations of each employee were obtained where possible. All immediate supervisors provided the ratings ($n = 5$); they were different managers from those who conducted the interviews. They were trained on the administrative purpose for the ratings, confidentiality of the information, definitions of the items, and importance of making distinctions among employees. Supervisors evaluated all employees in their areas who applied for the jobs.

Total scores were the average across the six items and across the supervisors. Because of mean differences between supervisors, their ratings were converted to z scores before averaging. Coefficient alpha reliability was .92, interrater reliability of the mean of the supervisors was .81, and average absolute interrater differences was .55 SD. Thus, the performance measure had high reliability and agreement.

Procedures

Employees were scheduled to take the interview and tests during nonwork time. They were not scheduled after working a night shift or after

overtime. Typically, tests were administered immediately after the interview. Job performance measures were collected concurrently. Strict control of all data was maintained. Supervisors involved in the interviews and appraisals did not score the instruments or keep copies. They were also instructed not to discuss their judgments with either employees or other supervisors.

Results

Table 2 presents means, standard deviations, and intercorrelations. There was substantial range and variation and little indication of ceiling effects on all measures. Education correlated positively with the tests (.46) and interview (.39) but was unrelated to job performance; conversely, tenure was unrelated to the tests and interview but correlated positively with performance (.36).

The interview and test composite correlated .60, indicating substantial shared variance. Both showed large correlations with job performance. Uncorrected validities were .50 for the interview and .46 for the test composite. This difference was not significant, $t(67) = 0.43$, ns. Validities corrected for the interrater reliability of .81 on the job performance measure were .56 and .51, respectively. Regression analyses revealed that the interview had incremental validity beyond the test composite, incremental $R^2 = .08$, $F(1, 67) = 7.19$, $p < .05$, and the test composite had incremental validity beyond the interview, incremental $R^2 = .04$, $F(1, 67) = 3.89$, $p < .05$.

Supplementary analyses divided the job performance measure into two composites—the two cognitive-related items (coefficient alpha = .86) and the four noncognitive items (.91)—and correlated them with the predictors (Table 2). For noncognitive performance, the interview correlated nonsignificantly higher than the tests, $rs = .42$ versus .35; $t(67) = 0.71$, ns. The interview had incremental validity beyond the tests, incremental $R^2 = .07$, $F(1, 67) = 5.69$, $p < .05$, but the tests did not have incremental validity beyond the interview, incremental $R^2 = .01$, $F(1, 67) = 1.22$, ns. Conversely, for cognitive performance, the tests correlated nonsignificantly higher than the interview, $rs = .60$ versus .57, $t(67) = 0.36$, ns, and the tests had incremental validity beyond the interview, incremental $R^2 = .11$, $F(1, 67) = 12.37$, $p < .05$. Yet, the interview also had incremental validity beyond the tests, incremental $R^2 = .06$, $F(1, 67) = 7.69$, $p < .05$.

Future questions had a slightly higher mean than past questions, Table 2; $t(69) = 2.67$, $p < .05$, but the variances were not different, $F(1, 68) = 2.50$, ns. The correlation between the future and past questions was .73. Coefficient alpha reliabilities of the 15-item future and past sections of the interview were .80 and .92, interrater reliabilities of the means were .94 and .97, and average absolute interrater differences were .19 and .21 scale points. Future questions had slightly lower correlations with cognitive abilities than did past questions (Table 2).

Table 2 shows past questions had higher validities than future questions, but the difference for total performance was not significant, .51 versus .39, $t(67) = 1.55$, ns. However, past questions had incremental validity beyond future questions in predicting total performance, incremental $R^2 = .11$, $F(1, 67) = 9.94$, $p < .05$, whereas future questions did not have incremental validity beyond past questions, incremental $R^2 = .00$, $F(1, 67) = 0.08$, ns. Finally, both question types had incremental validity be-

Table 2
Means, Standard Deviations, and Intercorrelations

Measure	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Total structured interview	3.65	0.52	—														
2. Future interview questions	3.72	0.45	90	—													
3. Past interview questions	3.57	0.67	95	73	—												
4. Test composite	0.00	0.81	60	49	61	—											
5. Reading comprehension	30.97	8.10	58	46	60	87	—										
6. Vocabulary	27.10	10.57	54	40	57	89	81	—									
7. Grammar	11.86	4.14	49	37	51	82	74	78	—								
8. Arithmetic	34.37	12.07	50	42	49	85	74	71	64	—							
9. Mathematical problem solving	7.56	3.72	54	44	55	86	70	73	69	78	—						
10. Measurement scales	8.23	5.10	45	36	46	79	61	65	56	60	63	—					
11. Mechanical reasoning	30.60	7.10	56	44	57	75	69	64	54	55	57	58	—				
12. Visual speed and accuracy	79.01	27.51	29	28	26	73	50	62	55	63	64	50	34	—			
13. Visual pursuit	14.08	6.03	44	38	42	73	54	57	45	55	55	59	56	53	—		
14. Total job performance	−0.06	0.73	50	39	51	46	47	42	37	36	37	38	36	22	41	—	
15. Cognitive job performance	−0.06	0.80	57	44	59	60	56	57	55	43	49	46	46	34	50	89	—
16. Noncognitive job performance	−0.05	0.77	42	33	43	35	38	31	24	30	28	30	27	14_a	33	97	75

Note. $n = 70$. All correlations significant at $p < .05$, except one marked with $_a$. Decimal points are omitted in the correlations.

yond the tests, incremental R^2s = .08 and .04, $F(1, 67) = 8.12$ and 3.30, $p < .05$ and .10, for past and future questions, and vice versa, incremental R^2s = .04 and .10, Fs$(1, 67) = 3.40$ and 8.56, ps $< .10$ and .05, respectively.

Discussion

The first purpose was to examine incremental validity of a structured interview beyond cognitive ability tests. The interview showed substantial correlations with the tests, but it still had meaningful incremental validity in predicting both cognitive and noncognitive performance criteria. The variety of tests used suggests they were not disadvantaged in this comparison. That both the interview and tests had incremental validity in most analyses indicates that both may be usefully included in the same selection system.

The interview was designed to complement the tests in the KSAOs it assessed, so it is not possible to attribute the results clearly to either measurement approach (interview versus tests) or content (different KSAOs). However, it is likely that measurement would not explain the results because highly structured interviews are not more standardized than written tests. Also, interviews assessing cognitive abilities have not shown incremental validity (Campion et al., 1988; Walters et al., 1993). Thus, the incremental validity was likely due to the different constructs assessed.

The second purpose was to examine whether past or future-oriented questions have higher validity. The past questions had higher validity, but not significantly higher probably due to modest statistical power (i.e., 67% to detect a "medium" difference; Cohen, 1977). However, they did have incremental validity beyond future questions, while future questions did not have incremental validity beyond past questions. Both question types had very comparable reliabilities and variances. Past questions had a slightly higher coefficient alpha, but correcting the validities did little to diminish the difference (corrected rs = .53 and .44). Past questions may have been slightly more diffi-

cult based on the lower mean, but future questions had more linkages in the content validity analysis. The past questions had slightly higher correlations with the cognitive tests, but it is unclear if this is an advantage because all the procedures had high validity. Thus, there are no obvious methodological reasons for the findings.

Theoretically, this may mean that past behavior is a slightly better predictor of future behavior than are future intentions. But three methodological unknowns qualify this explanation. First, the highly experienced candidates in this setting may make past questions more relevant. With inexperienced candidates (e.g., women, minorities, youths), there might be restriction of range with past questions and so future questions might be more valid. Second, future questions may be inhibited by fakability in selection contexts. Third, the highly structured format of situational interviews was used with both question types. It is unknown if relative effectiveness of the questions would be the same with the more flexible structure of behavior description interviews. In conclusion, it must be noted that both question types had high validity, both had incremental validity beyond cognitive tests, and the inclusion of both types may provide flexibility in item writing for test developers and variety for candidates and interviewers.

References

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35,* 281–322.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1991). *Technical manual for the differential aptitude tests.* Chicago: Psychological Corporation.

Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41,* 25–42.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Flanagan, J. C. (1975). *Flanagan industrial tests examiner's manual.* Chicago: London House.

Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology, 42,* 691–726.

Hunter, J. E., & Hunter, R. F. (1984). The validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98.

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

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., & Latham, G. P. (1984). *Goal-setting: A motivational technique that works.* Englewood Cliffs, NJ: Prentice-Hall.

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.

Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology, 70,* 774–776.

Pursell, E. D., Campion, M. A., & Gaylord, S. R. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal, 59,* 907–912.

Ramsay, R. T. (1991). *Manual for the job skills test—Reading.* Pittsburgh, PA: Author.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1–62.

Ruch, F. L., & Ruch, W. W. (1980). *Technical report for the employee aptitude survey.* Glendale, CA: Psychological Services.

Ruch, W. W., Weiner, J. A., McKillip, R. H., & Dye, D. A. (1985). *Technical manual for the PSI basic skills test.* Glendale, CA: Psychological Services.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology, 29,* 79–101.

Science Research Associates. (1986). *Reading and arithmetic indexes examiner's manual.* Chicago: London House.

Walters, L. C., Miller, M. R., & Ree, M. J. (1993) Structured interviews for pilot selection: No incremental validity. *International Journal of Aviation Psychology, 3*(1), 25–38.

Weekley, J. A., & Gier, J. A. (1987). Reliability and validity of the situational interview for a sales position. *Journal of Applied Psychology, 72,* 484–487.

Wesman, A. G. (1980). *Personnel tests for industry—Verbal test manual.* Chicago: Psychological Corporation.

Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61,* 275–290.

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology, 62,* 191–199.