

STRUCTURED INTERVIEWING: RAISING THE PSYCHOMETRIC PROPERTIES OF THE EMPLOYMENT INTERVIEW

MICHAEL A. CAMPION
Krannert School of Management
Purdue University

ELLIOTT D. PURSELL
Human Resource Systems
New Bern, North Carolina

BARBARA K. BROWN
Krannert School of Management
Purdue University

A highly structured employment interviewing technique is proposed, which includes the following steps: (1) develop questions based on a job analysis, (2) ask the same questions of each candidate, (3) anchor the rating scales for scoring answers with examples and illustrations, (4) have an interview panel record and rate answers, (5) consistently administer the process to all candidates, and (6) give special attention to job relatedness, fairness, and documentation in accordance with testing guidelines. Examination of psychometric properties for hiring entry-level production employees ($n = 149$) reveals high interrater reliability ($r = .88$) and predictive validity (uncorrected $r = .34$, corrected $r = .56$), as well as evidence for test fairness and utility. The levels of these properties are comparable to those of a comparison battery of typical employment tests, and correlations with the tests suggest that the interview has a strong cognitive aptitude component. Potential explanations for the effectiveness of this structured interviewing technique are discussed.

The reliability and validity of the employment interview has been questioned throughout the history of industrial psychology, starting early in the century (e.g., Hollingworth, 1922; Scott, 1915) and in every review since (Arvey & J. Campion, 1982; Carlson, Thayer, Mayfield, & Peterson, 1971; J. Hunter & R. Hunter, 1984; Mayfield, 1964; Reilly & Chao, 1982; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949; Webster, 1964; O. Wright, 1969). Structuring the interview has often been proposed as a means of improvement, but the operationalization of structure has varied widely across studies, with at least three different forms observed. First, many studies developed semistructured interviews in that the process

Special thanks to Chris J. Berger, James E. Campion, Michael M. Harris, Judith L. Komaki, James S. Russell, and two anonymous reviewers for their comments on this manuscript.

Correspondence and requests for reprints should be addressed to Michael A. Campion, Krannert School of Management, Purdue University, West Lafayette, IN 47907.

was not completely specified but some form of assistance was provided in conducting the interview or evaluating responses. Yonge (1956) used an outline and rating scale. Barrett, Svetlik, and Prien (1967) had interviewers take notes during the interview and make ratings afterwards. Landy (1976) used dimensional rating scales and an interview panel. Mayfield, Brown, and Hamstra (1980) developed an interview guide and summary form with suggested questions. Zedeck, Tziner, and Middlestadt (1983) made evaluations on behaviorally anchored rating scales. Arvey, Miller, Gould, and Burch (1987) developed a job-analysis-based interview schedule. In addition, many other studies provided some form of structure for the interview, but the detail given was insufficient for understanding the nature or degree of the structure (e.g., Albrecht, Glaser, & Marks, 1964; Borman, 1982; Campbell, Prien, & Brailey, 1960; Freeman, Manson, Katzoff, & Pathman, 1942; Gardner & Williams, 1973; Ghiselli, 1966; Glaser, Schwarz, & Flanagan, 1958; Huse, 1962; Tubiana & Ben-Shakhar, 1982).

Second, some studies "patterned" the interview as suggested by McMurry (1947). With this approach, the interviewer did not have to ask the same questions of each candidate but, instead, selected from an array (or pattern) of questions. Maas (1965) improved the patterned interview by using behavioral expectation scales (Smith & Kendall, 1963). Janz (1982) and Orpen (1985) evaluated a further variant called the "patterned behavior description interview," which involved a critical incident job analysis (Flanagan, 1954), recorded responses, and rating scales. Finally, Schwab and Heneman (1969) and Heneman, Schwab, Huett, and Ford (1975) enhanced structure in a manner similar to patterning, in which interviewers could not deviate from a predetermined job-analysis-based interview format (i.e., application form).

Third, Latham, Saari, Pursell, and M. Campion (1980), Latham and Saari (1984), and recently, Weekley and Gier (1987) evaluated an approach called "situational" interviewing. This approach was more structured than previous efforts in that it used the same questions for each candidate, anchored rating scales, and an interview panel to record and evaluate responses. On the basis of the goal-setting assumption that intentions are related to behavior (Locke, 1968), candidates' responses as to what they would do in hypothetical job situations were hypothesized to be predictive of what they would actually do on the job. Situations were generated using critical incident job analyses (Flanagan, 1954).

Purpose of this Study

This study extends previous research in four ways. First, it presents a more highly structured interviewing technique than most other efforts. Previously this technique was only described in a nonresearch journal (Pursell,

M. Campion, & Gaylord, 1980) and in unpublished sources (M. Campion & Pursell, 1981; Pursell & Gaylord, 1976). This technique begins with the same research theme as Latham et al. (1980), but it extends the methodology to other question types in addition to situational questions. Also included are job knowledge, worker requirements, and job sample and simulation questions. Second, more so than with other approaches to interviewing, greater explicit attention is afforded to the guidelines on test development from both the professional perspective (*Validation Principles*, Society of Industrial and Organizational Psychology, Inc., 1987) and the legal perspective (*Uniform Guidelines*; Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978) in order to enhance the likelihood of validity and legal defensibility. Third, aside from presenting the usual psychometric quality indicators of interrater reliability and predictive validity, this study also presents evidence of test fairness and utility. Fourth, a battery of typical employment aptitude tests are examined to determine whether the psychometric properties of the interview can be raised to the level of these traditional selection devices and to explore the constructs measured by the interview.

Overview of the Structured Interviewing Technique

Both the *Uniform Guidelines* (p. 38296) and the *Validation Principles* (p. 1) consider interviews to require validation just as any other selection procedure. The aforementioned common belief in the lack of validity of the traditional unstructured interview, along with its inherent subjectivity and apparent susceptibility to bias, may make it particularly vulnerable to legal attack (Arvey, 1979). The proposed approach to interviewing attempts to reduce subjectivity and inconsistency by highly structuring the process with the following six steps.

Step 1: Develop questions based on a job analysis. Adequate job analysis for any selection procedure is not only encouraged by the *Uniform Guidelines* (pp. 38304-38306) and *Validation Principles* (pp. 5-6), but there is evidence of its importance both in court decisions (Kleiman & Faley, 1985) and in avoiding bias (Kesselman & Lopez, 1979). All questions must be clearly job related. Any method of job analysis can be used, as long as it includes a determination of knowledge, skills, abilities, and other requirements upon which to base interview questions. There should be a measure of importance of job tasks so that questions only assess prerequisites for performing critical work (*Uniform Guidelines*, p. 38302; *Validation Principles*, pp. 22-23). Additionally, questions should not generally be based on the requirements of higher-level jobs (*Uniform Guidelines*, p. 38298; *Validation Principles*, pp. 13-14) or on knowledge or skills the

employee will learn with brief training or experience on the job (*Uniform Guidelines*, p. 38298; *Validation Principles*, p. 22).

A variety of question types can be used, including situational questions as described above (Latham et al., 1980), questions on job knowledge that is related to performance, and job sample or simulation questions where possible. These latter questions can range from actually performing part of the job to mock-ups of job tasks, or to simply phrasing questions in terminology and examples from the job. It is important, however, that they assess requirements at the same complexity level as that needed on the job (*Uniform Guidelines*, pp. 38305-38306). Samples and simulations not only enhance content and face validity but, when properly developed, can exhibit criterion-related validity (J. Campion, 1972) and avoid bias (Brugnoli, J. Campion, & Basen, 1979; Schmidt, Greenthal, J. Hunter, Berner, & Seaton, 1977). Finally, other worker requirement questions also are included. They frequently involve questions on background (e.g., experience, education) or "willingness" questions (e.g., shift work, travel). They also serve as warm-up questions at the beginning of the interview, and as realistic job previews (Wanous, 1980).

Aside from the criteria already mentioned, questions should be reviewed to make certain they are accurate, complete, and unambiguous. Furthermore, they should be reviewed by independent job experts who are members of protected groups to check for any potential for bias or misinterpretation.

Step 2: Ask the same questions of each candidate. All candidates are asked the same questions. There is no prompting or follow-up questioning, although the questions can be repeated.

Step 3: Anchor the rating scales for scoring answers with examples and illustrations. A scoring system is developed for each question by generating examples or definitions for *good* (5), *marginal* (3), and *poor* (1) answers. One approach is to ask job experts for example answers they have actually heard that subsequently distinguished different levels of performers on the job (Latham et al., 1980). A simpler approach is to brainstorm potential answers with experts and personnel representatives familiar with the job and with interviewing comparable candidates. Often both approaches are used to generate potential answers. Either way, example answers must be scaled to the requirements of the job so that good answers do not far exceed the requirements, and poor answers are not so low that they do not help distinguish between candidates. Predetermined answer-rating scales enhance consistency across interviews and objectivity of judging candidate responses. Making the scoring system explicit is essential to justifying the content validity of assessment procedures (Sackett, 1987; *Validation Principles*, p. 24). In addition, developing example answers serves as an

evaluation of the questions; difficulty in generating answers suggests that restructuring or elimination of the questions may be warranted.

Step 4: Have an interview panel record and rate answers. Using an interview panel reduces the impact of idiosyncratic biases that single interviewers might introduce (*Validation Principles*, p. 12). If possible, the panel should consist of a subset of the job experts who helped analyze the job and develop the interview questions because they are most familiar with the job and the questions. Three members are typically used, including supervisors of the job to be filled and a personnel representative. It is advisable to use the same members for all interviews to enhance consistency. However, an excessively large number of interviews or other constraints (e.g., turnover) may make this infeasible. The panel assembles in advance to review job duties and requirements, questions and answers, the interview process, and ways to avoid rating errors that can bias evaluations (Latham, Wexley, & Pursell, 1975). Application forms are not reviewed prior to the interview to avoid influencing the evaluation process (cf. Dipboye, Fontenelle, & Garner, 1984; Tucker & Rowe, 1977). All members independently record and rate each candidate's answers during the actual interview.

Step 5: Consistently administer the process to all candidates. The same panel member should conduct all interviews and ask all questions. Panel members do not discuss questions, answers, or candidates between interviews in order to avoid potential bias from changing standards or comparisons among candidates. After all interviews are complete for a given job, any large discrepancies between ratings are discussed (Thornton & Byham, 1982; but cf. Sackett & Wilson, 1982). Memory decay for candidate answers is avoided by the extensive note taking of panel members. The ratings and items are averaged to ensure equal weighting because differential weighting schemes are generally not preferred (Einhorn & Hogarth, 1975; Wainer, 1976). Differences in importance between job requirements is addressed by the number of items assessing each requirement. Every attempt is made to conduct interviews in as nonstressful a manner as possible (e.g., introductions of panel members, comfortable interview setting, only one member asks all questions). Candidates are allowed to ask questions in a subsequent nonevaluation interview with a personnel representative.

Step 6: Give special attention to job relatedness, fairness, and documentation in accordance with testing guidelines. Consideration of the *Uniform Guidelines* and *Validation Principles* has been noted throughout the process. Components needing documentation include the job analysis and interview development procedure, candidate responses and scores, evidence for content or criterion-related validity, adverse impact analyses, and other aspects as appropriate.

Method

Setting and Sample

The structured interview was used for hiring entry-level labor-pool employees in a large pulp and paper mill located in the rural Southeast. The traditional unionized facility placed new employees in bottom-level jobs in the various union lines of progression throughout the mill. Interest was in selecting employees with the basic skills needed to perform any of these entry-level jobs.

Of the 243 applicants interviewed, 149 were hired. The hires were 37.6% minority and 20.1% female, which was representative of the community work force and likely future candidates (*Uniform Guidelines*, p. 38301; *Validation Principles*, p. 12). Age averaged 30.4 ($SD = 7.9$) and education averaged 12.2 ($SD = 2.1$) years. Adequate statistical power was ensured (*Validation Principles*, p. 8) by the fact that the 149 hires provided 90% power to detect an observed correlation of .24 (i.e., uncorrected for range restriction and criterion unreliability; $p < .05$, one-tailed test; Schmidt, J. Hunter, & Urry, 1976).

Structured Interview

Analyses of the labor pool revealed that a subset of 17 jobs were most frequently staffed by entry-level employees. Job analysis conferences were conducted with incumbents and supervisors for each job. Lists of duties and requirements were generated, and duties were evaluated in terms of importance and time spent. The jobs were highly similar in terms of 25 shared requirements (i.e., knowledge, skills, abilities, and other worker characteristics), which suggested support for a job family for selection purposes (Pearlman, 1980; *Uniform Guidelines*, p. 38304). Interview questions were developed to assess the requirements needed to perform the most important and time-consuming duties. Detailed content validity procedures and analyses were described in M. Campion and Pursell (1981).

The developmental procedures described above were followed, and a 20-item interview resulted. All the previously mentioned question types were included, with most questions representing a combination of the various types. Three examples provided below illustrate the range of questions.

- a. Job knowledge question assessed mechanical comprehension: "When putting a piece of machinery back together after repairing it, why would you clean all the parts first?"
 - (5) Particles of dust and dirt can cause wear on moving parts. Need to have parts clean to inspect for wear and damage.
 - (3) Parts will go together easier. Equipment will run better.

- (1) So it will all be clean. I don't know.
- b. Simulation question assessed low level reading ability: "Many of the jobs require the operation of a forklift. Please read this (90-word) forklift checkout procedure aloud."
 - (5) Reads fluently pronouncing all words accurately.
 - (3) Can read most words but hesitates.
 - (1) Reads with great difficulty.
- c. Worker characteristic or willingness question assessed fear of heights: "Some jobs require climbing ladders to a height of a five-story building and going out on a catwalk to work. Give us your feeling about performing a task such as this."
 - (5) Heights do not bother me. I have done similar work at heights in the past (and gives examples).
 - (3) I do not think I am afraid of heights. I know that this would have to be done as part of the job.
 - (1) I am afraid of heights. I would do it if absolutely necessary.

A cutting score of 4.0 was set for the interview based on a modified Angoff (1971) procedure. Job experts judged the minimum acceptable performance level on each item, and the average across the items was used as the cutting score. This helped ensure that the interview assessed requirements at a similar complexity level to the job (*Uniform Guidelines*, pp. 38305-38306).

Because of the magnitude of this hiring program, many interview panel members were needed. But in all cases, the three panel members included two supervisors familiar with the different entry-level jobs and a personnel representative. Interviews lasted approximately 30 minutes.

Performance Criterion

Behavioral observation scales (Latham & Wexley, 1977, 1981) were developed because more objective, behavioral, and job-analysis-based performance appraisals may be more legally defensible (Kleiman & Faley, 1985; *Uniform Guidelines*, p. 38300). This procedure began with a critical incident job analysis (Flanagan, 1954), which was separate from the analysis used to develop the interview. Content coverage was ensured by collecting 6,150 critical incidents from 100 supervisors describing the performance of 393 employees. Incidents were condensed into 75 items and coupled with 5-point frequency scales to constitute the behavioral observation instrument. Reproducibility was demonstrated by having an independent analyst reclassify a 9% sample of the critical incidents into the 75 items. Approximately 96% accuracy was observed. Total scores were calculated as sums of the items. Internal consistency reliability (Cronbach,

1951) was .99, indicating substantial homogeneity. Interrater reliability between independent supervisory evaluations over a one-month time period on a sample of 30 employees was .76 ($p < .05$).

All supervisors using the instrument were given a half-day training program on minimizing rating errors (Latham et al., 1975). Appraisals were conducted after six months of employment. To avoid potential bias, in no case were the same supervisors involved in both the interviews and performance appraisals (Schoorman, 1988; *Uniform Guidelines*, p. 38300; *Validation Principles*, p. 14).

Employment Aptitude Tests

Before starting work, all candidates hired were examined on an experimental battery of four typical paper-and-pencil employment tests, which were chosen to assess cognitive aptitudes suggested by the job analysis.

1. Mathematical aptitude was measured with the Flanagan Industrial Test—Arithmetic (Form A). Its 60 items measured ability to perform simple addition, subtraction, multiplication, and division of whole numbers (Flanagan, 1975). A time limit of 15 minutes was allowed, and the score was the number correct. The manual reported an alternative forms reliability of .79.

2. Mechanical aptitude was measured with the mechanical knowledge test of the SRA Mechanical Aptitudes battery (Form AH). Its 45 items measured ability to recognize a variety of common tools (Richardson, Bellows, Henry, & Company, Inc., 1947). It was timed at 10 minutes and scored in terms of number correct. The manual only reported internal consistency reliability, which was inappropriate for speeded tests (Nunnally, 1978).

3. Following oral instructions was measured with the Personnel Tests for Industry—Oral Directions Test (Form S). Its 16 items measured ability to follow oral instructions on a 15-minute audio tape (Langmuir, 1974). The score was a weighted total of number correct (39 maximum). The manual reported a mean corrected split-half reliability of .80 and a mean retest reliability of .84.

4. Reading scales was measured by the 12-item Can You Read a Scale? test, which assessed ability to read standard scales or rules (Lawshe, 1943). It was timed at four minutes and scored as number correct. No manual was available.

Results

All measures appeared to have substantial range and variation (Table 1). Internal consistency reliability of the structured interview was .72,

TABLE 1
Means, Standard Deviations, and Correlations
among the Measures

	<i>n</i>	<i>M</i>	<i>SD</i>	Correlations ^a					
				1	2	3	4	5	6 ^b
1. Structured interview	149	4.21	.39		40	70	52	66	49/56
2. Mathematical aptitude	140	53.95	7.18	27		38	58	51	33/38
3. Mechanical aptitude	140	27.16	8.25	54	27		45	74	45/52
4. Following oral instructions	140	32.72	4.74	37	53	31		63	40/46
5. Reading scales	140	6.95	3.77	50	44	64	54		44/51
6. Performance appraisal	149	278.30	53.42	34	25	32	30	32	

^aDecimals omitted. All significant at $p < .05$. Correlations below the diagonal are uncorrected; those above are corrected for range restriction. Corrections involving the interview are for direct restriction, while those not involving interview are for indirect restriction (Guilford, 1965, pp. 343 and 344, respectively).

^bCorrelations to left of slash (/) corrected for range restriction; those to right also corrected for criterion unreliability (Guilford, 1965, p. 487).

indicating some heterogeneity among the 20 items. Interrater reliability (intraclass correlation) among the three raters was quite high at .88. The reliability of the mean of the three raters was very high at .96 (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Interrater agreement was also examined (Tinsley & Weiss, 1975) because hiring decisions were made based on the absolute level of interview scores. Agreement among interviewers was within a criterion of .5 points on the total scores in 95% of the cases ($p < .05$; Lawlis & Lu, 1972).

The validity coefficient between the structured interview and performance appraisal was .34 ($p < .05$). Correcting for direct range restriction caused by selection based on the interview (restricted *SD* in the sample hired of .39 versus unrestricted *SD* in the total sample interviewed of .60) and criterion unreliability (interrater $r = .76$) yielded correlations of .49 and .39, respectively (Guilford, 1965, pp. 343 and 487). Correcting for both factors yielded a correlation of .56 (Schmidt et al., 1976).

Test fairness was evaluated using a moderated regression strategy, which assessed equality of intercepts and slopes (Bartlett, Bobko, Mosier, & Hannan, 1978). In the equation in which the interview is used to predict the performance appraisal, intercept differences were tested by adding race to the equation, and the slope differences were tested by also adding the race by interview interaction to the equation. A similar analysis was performed for sex. Results showed a significant intercept difference for race (incremental *R* squared = .11, $F = 20.30$, $df = 1,146$, $p < .05$), but a plot of the separate regression lines indicated that a common line slightly over-predicted (i.e., was not unfair) for minorities. No slope differences were

observed for race, and no intercept or slope differences were observed for sex ($p > .05$).

Gain in utility from using the structured interview over random selection (*Validation Principles*, pp. 17-18) was estimated using formulas and procedures from Schmidt, J. Hunter, McKenzie, and Muldrow (1979). Relevant data included interviewer time and administrative costs of \$30 per applicant, selection ratio of .62, average standard score on the interview of those selected of .42, and validity coefficients of .34 uncorrected and .56 corrected. The standard deviation of job performance in dollar terms was estimated at \$5,000 per year by supervisors using the Schmidt et al. direct estimate technique. This value was 33% of annual mean wages, which was slightly below the 40% estimate often discussed in utility research (Schmidt & J. Hunter, 1983; Schmidt, J. Hunter, Outerbridge, & Trattner, 1986). Using these figures, the one-year utility from the 149 hires was estimated at approximately \$100,000 using the uncorrected validity and \$168,000 using the corrected validity. Assuming a 10% annual interest rate and no separations (Boudreau & Berger, 1985), the estimated gain in 10 years would be over \$1 million in net present value in the year 1980. Precise development costs were unknown, but they would be small compared with this gain in utility (e.g., \$20,000 to \$30,000 in salaries).

The four employment aptitude tests were also positively correlated with the performance appraisal (Table 1). Correcting for indirect range restriction caused by the interview substantially increased their size (Guilford, 1965, p. 344). Although the correlations were slightly smaller than the interview, none were significantly smaller ($p > .05$).

Test fairness analyses of the tests resulted in findings similar to the interview with regard to race. No slope differences occurred, but intercept differences were significant in all cases (incremental R squared = .07 to .11, $F = 12.02$ to 18.46 , $df = 1,137$, $p < .05$). Again, plots of the separate regression lines indicated that a common line slightly overpredicted (i.e., was not unfair) for minorities. With regard to sex, the tests showed no slope differences, but intercept differences were significant for the mechanical aptitude and following oral directions tests (incremental R squared = .06 and .05, respectively, $F = 9.86$ and 8.28 , $df = 1,137$, $p < .05$). Plots of regression lines indicated that the common line slightly overpredicted (i.e., was not unfair) for females.

Gain in utility from using the tests over random selection was analyzed assuming the same selection ratio, standard score for selectees, and standard deviation of job performance in dollars, but administrative costs of only \$5 per applicant. The one-year gain in utility ranged from approximately \$77,000 to \$100,000 using the uncorrected validities and from \$118,000 to \$160,000 using the corrected validities. These values were slightly smaller than those with the interview. Therefore, even though the tests would

have reduced development costs (i.e., job analysis still required, but no instrument development costs), the tests and interview would likely have comparable utility.

An examination of incremental value revealed that the tests explained additional variance in the performance appraisal beyond that explained by the interview (incremental R squared = .07, $F = 11.17$, $df = 4,135$, $p < .05$), but the interview did not explain additional variance in the appraisal beyond that explained by the tests combined (incremental R squared = .01, $F = 2.30$, $df = 1,135$, n.s.). Consequently, the interview would not have incremental utility beyond the tests, and there may even be a slight negative utility because of the development costs of the interview.

Correlations between the interview and tests were positive and moderate to large in size (Table 1), with a multiple correlation of .59 ($F = 17.99$, $df = 4,135$, $p < .05$). Correcting for range restriction on the interview substantially increased their size and resulted in a multiple correlation of .75. This suggested a strong cognitive aptitude component to the interview.

Discussion

This article presents a highly structured interviewing technique that includes the following steps: (1) develop questions based on a job analysis, (2) ask the same questions of each candidate, (3) anchor the rating scales for scoring answers with examples and illustrations, (4) have an interview panel record and rate answers, (5) consistently administer the process to all candidates, and (6) give special attention to job relatedness, fairness, and documentation in accordance with testing guidelines. In a field study, an interview developed using this technique demonstrates interrater reliability, predictive validity, test fairness for minorities and females, and cost/benefit utility.

Paper-and-pencil cognitive aptitude tests have historically been viewed as the best predictors of job performance (J. Hunter & R. Hunter, 1984; Reilly & Chao, 1982). Four typical employment tests are used to determine whether highly structuring the interview can raise its psychometric properties to the levels of these traditional selection instruments. The level of reliability of the interview seems comparable to those of the tests, to the extent that reliabilities for the tests are available in the manuals. The level of validity of the interview is slightly larger than the tests, but not significantly so. The corrected interview validity of .56 is also quite similar to the mean validity of .53 for cognitive aptitude predictors in general, and far larger than the mean validity of .14 for traditional interviews, for entry-level jobs as discovered in a large-scale meta-analytic study (J. Hunter & R. Hunter, 1984). Test fairness analyses yield similar results for both the interview and tests with regard to race, with a slight overprediction for

minorities. While the interview shows no difference for sex, however, the tests slightly overpredict females. Intercept differences are a frequent form of differential prediction (Bartlett et al., 1978), but slight overprediction is not unfair to minorities and females. Finally, the utilities of the interview and the tests are quite comparable, even with the larger development costs of the interview. Using expectancy tables (Taylor & Russell, 1939) and assuming base rates and selection ratios of 50%, the traditional interview ($r = .14$) yields 55% successful employees, and this structured interview ($r = .56$) yields nearly 70% successful employees.

The effectiveness of this approach to structured interviewing may be explained in terms of both method and content. Both clearly represent characteristics that are central to professional and legal testing guidelines. The methods used to structure the interview give it an advantage in terms of standardization (*Uniform Guidelines*, p. 38298; *Validation Principles*, pp. 7 and 14). Having multiple interviewers consistently evaluate candidates on the same questions using the same criteria may reduce idiosyncratic biases of interviewers, as well as their susceptibility to order or contrast effects among the candidates. Other methodological features may also enhance validity. All the question types have some prior empirical validity evidence: situational (e.g., Latham et al., 1980), knowledge (e.g., J. Hunter, 1986), sample/simulation (e.g., J. Campion, 1972), and worker requirements (e.g., Wanous, 1980). Content validity may be enhanced by attempting to assess requirements at the same complexity level as needed on the job (*Uniform Guidelines*, p. 38305–38306) and by explicitly predetermining the scoring system (Sackett, 1987). Training interviewers to avoid common response errors (Latham et al., 1975) may also help improve validity (Pursell, Dossett, & Latham, 1980).

The content of the structured interview may make it effective in that it is based on job analysis (*Uniform Guidelines*, pp. 38304–38306; *Validation Principles*, pp. 5–6). Examination of a battery of typical employment aptitude tests bears on the content of the interview by providing an assessment of the constructs it taps. The large correlations suggest that the interview has a substantial job knowledge or cognitive ability component. Thus, the content of the structured interview may be more like that of an orally administered cognitive ability test. The predictability of job performance through measures of job knowledge or general cognitive ability is well documented (Gottfredson, 1986; J. Hunter, 1986).

The superiority of structured interviews over unstructured interviews and the potential method versus content explanations for that superiority have also been noted by recent unpublished meta-analytic studies (McDaniel et al., 1987; Wiesner & Cronshaw, 1987; P. Wright, Lichtenfels, & Pursell, 1987). It is important to attend to the constructs underlying

interviews in meta-analyses. Lumping cognitive-oriented with motivation-oriented interviews would be misleading.

Further analyses of the employment aptitude tests reveal that they also predict the performance criterion, and they add incrementally to the prediction beyond the interview but not the reverse. This raises the question of why spend the additional effort and expense of using a structured interview, when commercially available and inexpensive paper-and-pencil ability tests predict just as well. At least three explanations can be offered. First, many (if not most) managers believe in the value of interviews, and they will frequently conduct interviews and allow them to influence hiring decisions, even when other more valid selection devices are available. In fact, there is no evidence that the continual warnings of researchers over the last 40 years about the limitations of the traditional interview have decreased its prevalence. Structured interviewing allows managers to take part in the selection process in an interviewer role, yet it gives the usually haphazard interview psychometric qualities comparable to a cognitive ability test.

Second, the development procedures of the structured interview may make it easier to content validate than commercially available written tests, which makes the technique especially appealing to small employers (Robinson, 1981). Third, there is some evidence that content-oriented job sample tests are perceived by both minority and majority applicants as fairer and more appropriate than written tests (Schmidt et al., 1977). Thus, the structured interviewing approach in this study may have greater face validity to applicants than commercially available aptitude tests.

At least four directions for future research can be suggested. First, future research could explore the usefulness of this structured interviewing technique for other jobs and settings. Both this study and research on the highly similar situational interview (Latham et al., 1980; Latham & Saari, 1984) have focused on jobs in the forest products industry. With the exception of small samples of foremen and hourly workers in Latham et al. and a small sample of sales people in the recent study by Weekley and Gier (1987), most data have come from entry-level hiring. Although the authors are aware of many other applications with different and higher-level jobs, further reliability and validity evidence is needed. It may be that higher-level jobs (e.g., management) would require some probing and follow-up questioning. Perhaps coding schemes for recording probes can be borrowed from survey interviewing research (e.g., Survey Research Center's Interviewing Manual, 1976). In addition, employment interviewing research might benefit by consideration of other findings in survey interviewing research, such as the importance of interviewer-interviewee interactions and nonverbal behavior (e.g., Beed & Stimson, 1985; Kahn & Cannell, 1957; Warwick & Lininger, 1975).

Second, future research could examine the relative effectiveness of the various types of questions. In the present study, most questions represented combinations of situational questions, job knowledge questions, sample/simulation questions, and worker requirement questions. But research on purer forms, such as the Latham et al. (1980) study on situational questions, could be undertaken.

Third, a potential advantage of the proposed form of structured interviewing is management acceptance. This is not the case if selection programs are too cumbersome (Mayfield et al., 1980). The experience of the authors is that managers appreciate the obvious fairness of asking the exact same questions of all candidates and having predetermined answer-rating scales. Participation in the interview panel is an interesting and involving experience for the managers. They feel the process makes an otherwise subjective and "soft" interview seem more objective and worthwhile. Although this structured interview may be somewhat constraining relative to traditional interviews (e.g., no follow-up questions, one member asking all questions), there is an unpublished study suggesting managers and attorneys view structured interviews as more practical and defensible (Latham & Finnegan, 1987). That study also found that employee hires did not differ in preferences for types of interviews, but student applicants believed the unstructured interview was advantageous in winning a lawsuit. This latter finding, along with the enhanced fairness perceptions of content-oriented measures observed by Schmidt et al. (1977), suggests candidate reactions to structured interviewing may also be favorable for the company. Future research could further explore both managerial and candidate reactions to structured interviews.

Fourth, future research could examine underlying mechanisms. Is this highly structured interviewing technique effective because of its degree of standardization and other methodological advantages; is it effective because it is job related and taps general cognitive ability which is predictive of performance on most jobs; or are both mechanisms operating?

REFERENCES

- Albrecht PA, Glaser EM, Marks J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology*, 48, 351-360.
- Angoff WH. (1971). Scales, norms, and equivalent scores. In Thorndike RL (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Arvey RD. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, 86, 736-765.
- Arvey RD, Campion JE. (1982). The employment interview: A summary and review of recent research. *PERSONNEL PSYCHOLOGY*, 35, 281-322.
- Arvey RD, Miller HE, Gould R, Burch P. (1987). Interview validity for selecting sales clerks. *PERSONNEL PSYCHOLOGY*, 40, 1-12.

- Barrett GV, Svetlik B, Prien EP. (1967). Validity of the job-concept interview in an industrial setting. *Journal of Applied Psychology*, 51, 233-235.
- Bartlett CJ, Bobko P, Mosier SB, Hannan R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *PERSONNEL PSYCHOLOGY*, 31, 233-241.
- Beed TW, Stimson RJ (Eds.). (1985). *Survey interviewing: Theory and techniques*. Sydney, Australia: George Allen Unwin.
- Borman WC. (1982). Validity of behavioral assessment for predicting military recruiter performance. *Journal of Applied Psychology*, 67, 3-9.
- Boudreau JW, Berger CJ. (1985). Decision-theoretic utility analysis applied to employee separations and acquisitions [Monograph]. *Journal of Applied Psychology*, 70, 581-612.
- Brugnoti GA, Campion JE, Basen JA. (1979). Racial bias in the use of work samples for personnel selection. *Journal of Applied Psychology*, 64, 119-123.
- Campbell JT, Prien EP, Brailey LG. (1960). Predicting performance evaluations. *PERSONNEL PSYCHOLOGY*, 13, 435-440.
- Campion JE. (1972). Work sampling for personnel selection. *Journal of Applied Psychology*, 56, 40-44.
- Campion MA, Pursell ED. (1981). *Plymouth fiber extraboard validation report: Content and criterion-related validation of a structured interview and written test battery for an entry-level production job family*. Plymouth, NC: Human Resources Research, Weyerhaeuser Company.
- Carlson RE, Thayer PW, Mayfield EC, Peterson DA. (1971). Improvements in the selection interview. *Personnel Journal*, 50, 268-275, 317.
- Cronbach LJ. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Dipboye RL, Fontenelle GA, Garner K. (1984). Effects of previewing the application on interview process and outcomes. *Journal of Applied Psychology*, 69, 118-128.
- Einhorn HJ, Hogarth RM. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13, 171-192.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice. (1978). Adoption by four agencies of uniform guidelines on employee selection procedures. *Federal Register*, 43, 38290-38315.
- Flanagan JC. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Flanagan, JC. (1975). *Flanagan industrial tests examiner's manual*. Chicago: Science Research Associates.
- Freeman GL, Manson GE, Katzoff ET, Pathman JH. (1942). The stress interview. *Journal of Abnormal and Social Psychology*, 37, 427-447.
- Gardner KE, Williams APO. (1973). A twenty-five year follow-up of an extended interview selection procedure in the Royal Navy. *Occupational Psychology*, 47, 1-13.
- Ghiselli EE. (1966). The validity of a personnel interview. *PERSONNEL PSYCHOLOGY*, 19, 389-394.
- Glaser R, Schwarz PA, Flanagan JC. (1958). The contribution of interview and situational performance procedures to the selection of supervisory personnel. *Journal of Applied Psychology*, 42, 69-73.
- Gottfredson LS. (1986). The g factor in employment [Special issue]. *Journal of Vocational Behavior*, 29(3).
- Guilford JP. (1965). *Fundamental statistics in psychology and education* (4th ed.). New York: McGraw-Hill.

- Heneman HG, Schwab DP, Huett DL, Ford JJ. (1975). Interviewer validity as a function of interview structure, biographical data, and interviewee order. *Journal of Applied Psychology*, 60, 748-753.
- Hollingworth HL. (1922). *Judging human character*. New York: Appleton.
- Hunter JE. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter JE, Hunter RF. (1984). The validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Huse EF. (1962). Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. *PERSONNEL PSYCHOLOGY*, 15, 195-205.
- Janz T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577-580.
- Kahn RL, Cannell CF. (1957). *The dynamics of interviewing*. New York: Wiley.
- Kesselman GA, Lopez FE. (1979). The impact of job analysis on employment test validation for minority and nonminority accounting personnel. *PERSONNEL PSYCHOLOGY*, 32, 91-108.
- Kleiman LS, Faley RH. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. *PERSONNEL PSYCHOLOGY*, 38, 803-833.
- Landy FL. (1976). The validity of the interview in police officer selection. *Journal of Applied Psychology*, 61, 193-198.
- Langmuir CR. (1974). *Personnel tests for industry—Oral directions test manual*. New York: Psychological Corporation.
- Latham GP, Finnegan BJ. (1987). *The practicality of the situational interview*. Unpublished manuscript.
- Latham GP, Saari LM. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology*, 69, 569-573.
- Latham GP, Saari LM, Pursell ED, Campion MA. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Latham GP, Wexley KN. (1977). Behavioral observation scales for performance appraisal purposes. *PERSONNEL PSYCHOLOGY*, 30, 255-268.
- Latham GP, Wexley KN. (1981). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.
- Latham GP, Wexley KN, Pursell ED. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, 60, 550-555.
- Lawlis GI, Lu E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin*, 78, 17-20.
- Lawshe CF. (1943). Can you read a scale? In Tiffin J (Ed.), *Purdue vocational series*. West Lafayette, IN: University Book Store.
- Locke EA. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3, 157-189.
- Maas JB. (1965). Patterned scaled expectation interview: Reliability studies on a new technique. *Journal of Applied Psychology*, 49, 431-433.
- Mayfield EC. (1964). The selection interview—A re-evaluation of published research. *PERSONNEL PSYCHOLOGY*, 17, 239-260.
- Mayfield EC, Brown SH, Hamstra BW. (1980). Selection interviewing in the life insurance industry: An update of research and practice. *PERSONNEL PSYCHOLOGY*, 33, 725-739.
- McDaniel MA, Whetzel DL, Schmidt FL, Hunter JE, Maurer S, Russell J. (1987). *The validity of employment interviews: A review and meta-analysis*. Manuscript submitted for publication.
- McMurry RN. (1947). Validating the patterned interview. *Personnel*, 23, 263-272.

- Nunnally JC. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Orpen C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology*, 70, 774-776.
- Pearlman K. (1980). Job families: A review and discussion of their implications for personnel selection. *Psychological Bulletin*, 87, 1-28.
- Pursell ED, Campion MA, Gaylord SR. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal*, 59, 907-912.
- Pursell ED, Dossett DL, Latham GP. (1980). Obtaining valid predictors by minimizing rating errors in the criterion. *PERSONNEL PSYCHOLOGY*, 33, 91-96.
- Pursell ED, Gaylord SR. (1976). *Structured interviewing*. Tacoma, WA: Organizational Effectiveness, Weyerhaeuser Company.
- Reilly RR, Chao GT. (1982). Validity and fairness of some alternative employee selection procedures. *PERSONNEL PSYCHOLOGY*, 35, 1-62.
- Richardson, Bellows, Henry, & Company, Inc. (1947). *SRA mechanical aptitudes*. Chicago: Science Research Associates.
- Robinson DD. (1981). Content-oriented personnel selection in a small business setting. *PERSONNEL PSYCHOLOGY*, 34, 77-87.
- Sackett PR. (1987). Assessment centers and content validity: Some neglected issues. *PERSONNEL PSYCHOLOGY*, 40, 13-25.
- Sackett PR, Wilson MA. (1982). Factors affecting the consensus judgment process in managerial assessment centers. *Journal of Applied Psychology*, 67, 10-17.
- Schmidt FL, Greenthal AL, Hunter JE, Berner JG, Seaton FW. (1977). Job sample vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. *PERSONNEL PSYCHOLOGY*, 30, 187-197.
- Schmidt FL, Hunter JE. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedures utility. *Journal of Applied Psychology*, 68, 407-414.
- Schmidt FL, Hunter JE, McKenzie RS, Muldrow TW. (1979). The impact of valid selection procedures on work force productivity. *Journal of Applied Psychology*, 64, 609-626.
- Schmidt FL, Hunter JE, Outerbridge AN, Trattner MH. (1986). The economic impact of job selection methods on size, productivity, and payroll costs of the federal work force: An empirically based demonstration. *PERSONNEL PSYCHOLOGY*, 39, 1-29.
- Schmidt FL, Hunter JE, Urry VW. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology*, 61, 473-485.
- Schmitt N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *PERSONNEL PSYCHOLOGY*, 29, 79-101.
- Schoorman FD. (1988). The escalation bias in performance appraisal: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology*, 73.
- Schwab DP, Heneman HG. (1969). Relationship between interview structure and interinterviewer reliability in an employment situation. *Journal of Applied Psychology*, 53, 214-217.
- Scott WD. (1915). The scientific selection of salesmen. *Advertising and Selling*, 25, 5-6, 94-96.
- Smith PC, Kendall L. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149-155.
- Society of Industrial and Organizational Psychology, Inc. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.
- Survey Research Center, University of Michigan. (1976). *Interviewer's manual* (rev. ed.). Ann Arbor, MI: Institute for Social Research.

- Taylor HC, Russell JT. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Thornton GC, Byham WC. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Tinsley HEA, Weiss DJ. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22, 358-376.
- Tubiana JH, Ben-Shakhar G. (1982). An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. *PERSONNEL PSYCHOLOGY*, 35, 349-357.
- Tucker DH, Rowe PM. (1977). Consulting the application form prior to the interview: An essential step in the selection process. *Journal of Applied Psychology*, 62, 283-287.
- Ulrich L, Trumbo D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100-116.
- Wagner R. (1949). The employment interview: A critical summary. *PERSONNEL PSYCHOLOGY*, 2, 17-46.
- Wainer H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213-217.
- Wanous JP. (1980). *Organizational entry: Recruitment, selection, and socialization of newcomers*. Reading, MA: Addison-Wesley.
- Warwick DP, Lininger CA. (1975). *The sample survey: Theory and practice*. New York: McGraw-Hill.
- Webster EC. (1964). *Decision making in the employment interview*. Montreal: Eagle.
- Weekley JA, Gier JA. (1987). Reliability and validity of the situational interview for a sales position. *Journal of Applied Psychology*, 72, 484-487.
- Wiesner WH, Cronshaw SF. (1987). *The moderating impact of interview format and degree of structure on interview validity*. Manuscript submitted for publication.
- Wright OR. (1969). Summary of research on the selection interview since 1964. *PERSONNEL PSYCHOLOGY*, 22, 391-413.
- Wright PM, Lichtenfels PA, Pursell ED. (1987). *The structured interview: Additional studies and a meta-analysis*. Manuscript submitted for publication.
- Yonge KA. (1956). The value of the interview: An orientation and a pilot study. *Journal of Applied Psychology*, 40, 25-31.
- Zedeck S, Tziner A, Middlestadt SE. (1983). Interviewer validity and reliability: An individual analysis approach. *PERSONNEL PSYCHOLOGY*, 36, 355-370.