Krupa, D.J., Thompson, J.K., & Thompson, R.F. (1993). Localization of a memory trace in the mammalian brain. *Science, 260,* 989–991.

McCormick, D.A., & Thompson, R.F. (1984). Neuronal responses of the rabbit cerebellum during acquisition and performance of a classically conditioned nictitating membrane-eyelid response. *Journal of Neuroscience, 4,* 2811–2822.

Romano, A.G., & Patterson, M.M. (1987). The rabbit in Pavlovian conditioning. In I. Gormezano, W.F. Prokasy, & R.F. Thompson (Eds.), *Classical conditioning* (3rd ed., pp. 1–36). Hillsdale, NJ: Erlbaum.

Steinmetz, J.E. (1996). The brain substrates of classical eyeblink conditioning in rabbits. In J.R. Bloedel, T.J. Ebner, & S.P. Wise (Eds.), *The acquisition of motor behavior in vertebrates* (pp. 89–114). Cambridge, MA: MIT Press.

Steinmetz, J.E., Lavond, D.G., & Thompson, R.F. (1989). Classical conditioning in rabbit using pontine nucleus stimulation as a conditioned stimulus and inferior olive stimulation as an unconditioned stimulus. *Synapse, 3,* 225–233.

Steinmetz, J.E., Logue, S.F., & Steinmetz, S.S. (1992). Rabbit classically conditioned eyelid responses do not reappear after interpositus nucleus lesion and extensive post-lesion training. *Behavioural Brain Research, 51,* 103–114.

Steinmetz, J.E., & Sengelaub, D.R. (1992). Possible conditioned stimulus pathway for classical eyelid conditioning in rabbits: I. Anatomical evidence for direct projections for the pontine nuclei to the cerebellar interpositus nucleus. *Behavioral and Neural Biology, 57,* 103–115.

Welsh, J.P., & Harvey, J.A. (1989). Cerebellar lesions and the nictitating membrane reflex: Performance deficits of the conditioned and unconditioned response. *Journal of Neuroscience, 9,* 299–311.

Structuring Employment Interviews to Improve Reliability, Validity, and Users' Reactions

Michael A. Campion, David K. Palmer, and James E. Campion[1]

Department of Management, Purdue University, West Lafayette, Indiana (M.A.C.); Department of Management and Marketing, University of Nebraska, Kearney, Nebraska (D.K.P.); and Department of Psychology, University of Houston, Houston, Texas (J.E.C.)

In the 80 years of published research on employment interviews, one of the most strongly supported conclusions is that structuring the interview enhances its reliability and validity and, hence, its usefulness for prediction and decision making. "Structure" is broadly defined as any enhancement that increases standardization or otherwise assists the interviewer in determining what questions to ask or how to evaluate the responses.

The purpose of this article is to briefly summarize a recent review of the research literature that described and evaluated the many ways interviews can be structured (for more details, see Campion, Palmer, & Campion, 1997). From a review of nearly 200 articles and books, we have identified 15 components of structure that may enhance either the content of the interview or the evaluation process in the interview. Seven components influence the content of the interview by seeking to improve the quality and quantity of the information elicited. Eight components influence the evaluation process by seeking to improve the quality of the assessment of the information gathered, thus leading to a better decision regarding the interviewee.

We describe each component briefly and then critique its impact on reliability, validity, and users' reactions (as summarized in Table 1). Reliability is concerned with the consistency and standardization with which information is elicited across candidates and interviewers, thus facilitating comparison. Six types of reliability were considered: (a) test-retest reliability, or the extent to which the interviewer is consistent across time; (b) interrater reliability, or the extent to which there are consistent patterns between the ratings of different interviewers; (c) candidate consistency, or the extent to which the interview elicits consistent responding from a given candidate and across candidates; (d) interviewer-candidate interactions, or the extent to which differences in interaction quality are minimized across candidates; (e) internal consistency, or the extent to which the interview items are sufficiently numerous and intercorrelated so that the composite measures a homogeneous construct; and (f) interrater agreement, or the extent to which interviewers' absolute ratings agree.

Validity is concerned with whether the interview measures what it is supposed to measure, which usually means whether it

Recommended Reading

Campion, M.A., Palmer, D.K., & Campion, J.E. (1997). (See References)

Dipboye, R.L. (1992). *Selection interviews: Process perspectives.* Cincinnati, OH: South-Western.

Gollub Williamson, L., Campion, J.E., Malos, S.B., Roehling, M.V., & Campion, M.A. (1997). The employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82,* 900–912.

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

predicts future job performance. Three aspects of validity were evaluated: (a) job relatedness, or the extent to which the interview is related to the content of the job; (b) reduced deficiency, or the extent to which the interview minimizes the omission of relevant information; and (c) reduced contamination, or the extent to which the interview prevents irrelevant information from entering the process.

Three types of users' reactions were evaluated: (a) reduced equal employment opportunity (EEO) bias, or the extent to which intentional and unintentional illegal discrimination are reduced and perceptions of fairness are enhanced; (b) candidates' reactions, or the extent to which candidates might view the interview positively; and (c) interviewers' reactions, or the extent to which interviewers might view the interview positively.

STRUCTURING THE CONTENT OF THE INTERVIEW

Base Questions on a Job Analysis

A job analysis is required by both professional and legal testing guidelines. Any of a wide variety of job-analysis methods can be used, but the critical-incidents approach is most common. Critical incidents are examples of job-related behavior that distinguish among outstanding, average, and poor performance. They can provide ideas for interesting and job-related interview questions.

Basing questions on a job analysis should enhance validity by increasing job relatedness and by ensuring that the interview neither includes irrelevant information nor excludes relevant information. A job analysis also reduces EEO bias, and interview participants should react positively to the job-related focus.

Ask the Same Questions of Each Candidate

The earliest recognized and most basic component of structure is the standardization of questioning. Varying degrees of standardization are possible. The most common approaches are either to ask every candidate the same questions in the same order with no deviation or to ask primarily the same questions but allow some discretion (e.g., pick from a list of questions or have a common core of questions plus discretionary questions).

This component standardizes the samples of behavior to be judged by the interviewer, thus enhancing several types of reliability. It also decreases potential contamination by irrelevant questions, and it decreases deficiency by preventing omitted questions. By ensuring that all candidates are treated the same, it may reduce EEO bias (and therefore increase the legal defensibility of employment procedures).

Limit Prompting, Follow-up Questioning, and Elaboration on Questions

Some people feel that dynamic two-way interaction is an essential feature of the interview, and follow-up questioning facilitates such interaction. However, prompts and follow-up questions are the primary means by which an interviewer might bias the information gathering. The interviewer can actually lead the candidate to the right (or wrong) answer by the follow-up questions asked. The highest form of structure would be to prohibit follow-up questioning and prompts or to allow only standardized ones (e.g., repeating questions, asking the candidate for further explanation). This component involves a trade-off because sometimes follow-up questions are needed to get full information, explore negative answers, test hypotheses about the candidate, and so on.

Because it enhances standardization, this component should enhance most types of reliability. Its influence on validity is uncertain because of the trade-off just noted. Likewise, although it ensures equal treatment, and therefore reduces EEO bias, it may not be popular with users because it limits their freedom.

Use Better Types of Questions

Question type refers to how the question is asked. Question types are not neatly ordered by degree of structure, yet several types of questions appear relatively structured because they are specific and are usually used in conjunction with other components of structure, such as basing questions on job analysis and asking every interviewee the same questions. These types of questions are questions that pose hypothetical situations, questions that require answers describing past behaviors, questions concerning the candidate's background (e.g., education, work experiences), and questions asking for specific demonstration of job knowledge.

Better questions may increase some types of reliability by enhancing consistency, but the most likely effect is increased validity through enhanced job relatedness and through reduced contamination from low-quality questions. Better questions may also logically enhance fairness and interviewers' reactions.

Table 1. *Effects of interview structure on reliability, validity, and users' reactions*

Component	Reliability						Validity			Users' reactions		
	Test-retest reliability	Interrater reliability	Candidate consistency	Interviewer-candidate interaction	Internal consistency	Interrater agreement	Job relatedness	Reduced deficiency	Reduced contamination	Reduced equal employment opportunity bias	Candidates' reactions	Interviewers' reactions
Content												
Job analysis							+	+	+	+	+	+
Same questions	+	+	+	+				+	+	+		
Limited prompting	+	+	+	+				−	+	+	−	−
Better questions			+		+		+		+	+		+
Appropriate interview length	+	+			+			+			−	−
Ancillary information controlled	+	+				+		−	+	+	−	−
No questions from candidate	+	+	+	+				−	+		−	−
Evaluation												
Each answer rated or multiple scales used	+	+			+			+	+			
Anchored rating scales	+	+				+	+	+	+	+		+
Detailed notes	+	+				+	+	+	+	+		−
Multiple interviewers	+	+		+	+	+		+	+	+	−	
Same interviewer(s) for all candidates	+			+					+	−		
No discussion between interviews	+	−				−			+	+		−
Training	+	+	+	+	−	+	+	+	+	+	+	+
Statistical prediction	+	+			+			+	+	+		

Note. "+" means positive effect and "−" means negative effect. Taken from Campion, Palmer, and Campion (1997, p. 657).

Use an Appropriate Interview Length and a Sufficient Number of Questions

This is a basic, but often overlooked, component of structure. In the studies we reviewed that reported data, the mean length of interviews was 39 min (SD = 25.8) and 16.5 questions (SD = 8.7). The vast majority of the interviews lasted 30 to 60 min and included 15 to 20 questions.

Using an appropriate interview length increases various forms of reliability, including test-retest reliability, interrater reliability, and internal consistency. It may increase validity by decreasing deficiency. However, overly long interviews could be viewed negatively by users, both candidates and interviewers.

Withhold or Control Ancillary Information

Examples of ancillary information include application forms, work histories, personnel files, test scores, letters of reference, transcripts, recommendations, and the results of previous interviews. These sources of information should be used in the final evaluation of a candidate, but their use by the interviewer may be a considerable threat to structure. Ancillary information confounds the interview with other selection information and may decrease reliability if different information is available for different candidates or is weighted differently by different interviewers.

Either withholding or standardizing this information should increase test-retest and interrater reliability, as well as interrater agreement. The effects on validity are uncertain, however. Withholding ancillary information may reduce contamination in some cases, but it could also inadvertently increase deficiency if the information is valid. Users' reactions may be mixed. EEO bias should be reduced. However, interviewers may react negatively to not having all available information, and candidates may perceive the interviewer as unprepared or uninterested and react negatively.

Do Not Allow Questions From the Candidate Until After the Interview

Allowing uncontrolled questions from the candidate is a generally unrecognized source of variability that can compromise reliability and validity across candidates. It reduces standardization by changing the interview content in unpredictable ways.

Not allowing questions from candidates should standardize the content, thus increasing several types of reliability. The effects on validity are mixed because although this restriction could decrease contamination, at the same time, it could increase deficiency if the candidates' questions increase valid information. Also, interviewers and candidates may react negatively to this restriction.

In practice, the interview serves a recruitment as well as selection function. Candidates use the interview as a forum for gathering information about the job and the organization to use in their own decision-making process. This function can be served if each candidate is allowed time outside the interview to ask questions.

STRUCTURING THE EVALUATION PROCESS

Rate Each Answer or Use Multiple Scales

There are two elements to this component. First, the interviewer can rate each answer or the entire interview. Second, the interviewer can use multiple scales (e.g., for rating dimensions such as creativity, decisiveness, and communication ability) or only a single global scale. Rating each answer is more structured than rating the entire interview because judgments are more linked to specific responses; multiple scales are more structured than a global scale because they are more extensive. Although both approaches enhance the interview, rating each answer is somewhat more structured than using multiple scales.

This component should increase several types of reliability because it enhances the consistency of evaluation. Rating each answer may also be less cognitively complex and reduce memory requirements, thus further enhancing reliability. Deficiency is reduced through the use of more evaluations, and contamination is reduced because rating answers to individual questions using specific dimensions keyed to those questions focuses evaluation on relevant behaviors. Thus, validity should increase. Users' reactions should not be affected by this component.

Use Detailed, Anchored Rating Scales

Typically, four types of anchors are used: (a) example answers or illustrations, (b) descriptions or definitions of elements to look for in answers, (c) evaluative adjectives (e.g., "excellent," "marginal"), and (d) relative comparisons (e.g., top 20% of all candidates). Higher levels of structure are expected with the use of more and better anchoring (i.e., anchors phrased in terms of job behaviors and requirements, as determined through a job analysis).

Anchored rating scales are presumed to enhance objectivity; thus, they are expected to increase several types of reliability. If objectiv-

ity increases accuracy, then such scales may also reduce contamination and deficiency. If anchors are constructed using appropriate behaviors ascertained from a job analysis, then job relatedness may be enhanced. Objectivity should enhance EEO defensibility because it reduces potential for bias, and interviewers may appreciate the help anchors afford in making difficult judgments.

Take Detailed Notes

Note taking may enhance structure because it places less emphasis on memory recall and avoids certain possible effects of memory on ratings. For example, sometimes memory is best for events that occur early or late in a sequence (e.g., an initial handshake upon meeting the interviewer, the last question of the interview), but ideally interview ratings should reflect information gathered throughout the interview. The benefits of note taking may be most apparent when ratings are made at the end of the interview or are based on multiple questions.

Because notes become a written record available for examination by other people, note taking requires justifying the ratings, and this encourages interviewers to attend to candidates' responses and to organize their thoughts, thus possibly increasing accuracy, recall, and evaluation consistency. This should improve several types of both reliability and validity. Users' reactions are mixed. For example, candidates (and interviewers) may find note taking distracting, but it demonstrates to the candidate that the interviewer is paying attention.

Use Multiple Interviewers

The use of multiple interviewers improves the interview in several ways, such as by reducing the effects of idiosyncratic biases, increasing recall of the candidate's responses, averaging out random errors, and helping interviewers become aware of improper inferences. Multiple interviewers can conduct interviews together (e.g., a panel, or board, interview), or they can conduct them separately (e.g., a serial interview). Panel interviews are somewhat more structured than serial interviews.

The use of multiple raters will increase a number of different reliabilities. Deficiency should be reduced because relevant information is less likely to be missed, and contamination should be reduced because interviewers provide a check on each other. Thus, validity should be higher. Using multiple interviewers may reduce EEO bias if they reduce the effects of idiosyncratic biases. However, candidates' reactions might be negative if panel interviews are too stressful.

Use the Same Interviewer for All Candidates

The range of structure of this component is from one person conducting all the interviews (most structured) to a different person conducting each interview (least structured). Using one interviewer is often impractical (e.g., because of cost, time, geography), and using a different interviewer for each candidate is unusual, so the level of structure is a matter of degree. This component is particularly important when other components are unstructured because different interviewers tend to ask different questions and evaluate answers differently.

Using fewer interviewers should enhance consistency and therefore reliability. It may also increase validity if it reduces contamination due to differences between interviewers. However, perceptions of EEO bias may increase if the use of fewer interviewers makes the process appear more idiosyncratic.

Do Not Discuss Candidates or Answers Between Interviews

Discussing candidates or answers between interviews may introduce irrelevant information into the decision process or cause the evaluation criteria to change from one interview to the next.

Reliability effects may be mixed, but this component could enhance validity if it reduces contamination. It may also reduce perceptions of EEO bias if it prevents consideration of irrelevant information, but interviewers may react negatively to restrictions on their freedom.

Provide Extensive Interview Training

Training is probably the most common step taken by organizations to improve their selection interviews. It is less a component of structure per se than a way to ensure that other components are implemented correctly. Interviewing is relatively easily taught. The content of typical training programs includes establishing rapport, understanding the job requirements, asking questions and probing, evaluating answers, avoiding rating errors, ensuring equal employment opportunity, and other topics. The training techniques typically include lecturing, modeling, practicing and role playing, discussing, and videotaping.

Because training is intended to ensure the proper implementation of the other components, it should have a positive effect on most types of reliability, validity, and user reactions. Interestingly, training may have a negative effect on internal consistency if interviewers, when making their ratings, add in more variability than is warranted in a misguided attempt to correct for potential errors.

Use Statistical Rather Than Clinical Prediction

Different interviewers weigh information differently. Therefore, another way to enhance structure is to use statistical procedures rather than interview judgments to combine data. The most common approach is to simply sum or average the ratings. Within the context of employment interviews, statistical procedures may involve combining ratings across questions or dimensions, combining ratings across interviewers, and combining information from the interview with other information (e.g., test scores) to make final decisions.

Because it ensures consistency, statistical prediction should enhance several types of reliability. Validity may be improved if these procedures reduce the likelihood that contaminating information will enter the score or if they reduce the likelihood that relevant information will be overlooked. The potential for EEO bias should be reduced, but other reactions of users should be unaffected.

WHICH COMPONENTS ARE MOST IMPORTANT?

The most beneficial components of structure to use will vary depending on the nature of the context, the candidates, and the interviewers. Given that an organization has limited resources, which of the components would it be most beneficial for the organization to maximize? Basing our judgment on previous research, and conceptual arguments as to which components may most influence reliability and validity in a positive manner, thus affording the greatest potential gain over current typical practices, we suggest the following as the most influential components:

- Base questions on a job analysis
- Ask the same questions of each candidate
- Use better types of questions
- Rate each answer or use multiple scales
- Use detailed, anchored rating scales
- Provide extensive interview training

The benefits of structure may be offset by users' negative reactions to some of the components. However, there is very little research evidence at present to predict users' reactions with certainty. It is also possible that users will appreciate the obvious fairness and job relatedness of structured interviews, or they may appreciate how such interviews can help interviewers make difficult interview decisions. If users react adversely to a high level of structure, perhaps a more moderate level will be acceptable while still improving the quality of the interview.

CONCLUSIONS

Our conclusions can be summarized as follows: First, there is clear evidence for the superiority of structured interviews over unstructured interviews. Second, interviews can be structured in many different ways. Third, the components of structure are easy to implement, so there seems to be little reason not to structure interviews used for employment selection.

Note

1. Address correspondence to Michael A. Campion, 1310 Krannert Bldg., West Lafayette, IN 47907; e-mail: campionm@mgmt.purdue.edu.

Reference

Campion, M.A., Palmer, D.K., & Campion, J.E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50,* 655–702.