Why Don't You Just Show Me? Performance interviews for skill-based promotions

Frederick P. Morgeson*, Michael A. Campion and Julia Levashina*****

*The Eli Broad Graduate School of Management, Michigan State University, N475 North Business Complex, East Lansing, MI 48824 1122, USA. morgeson@msu.edu
**Krannert School of Management, Purdue University, West Lafayette, IN 47907, USA
***College of Business Administration, Kent State University, PO Box 5190, Kent, OH 44242-0001, USA

Despite the high skill levels needed to cope with complex technical systems and the pace of technological change, there remain persistent skill gaps in the United States workforce. Organizations are increasingly relying on skill-based programs to encourage and foster employee skill development. Unfortunately, many questions remain about how to make skill-based promotions. Drawing from research on performance testing and structured interviewing, as well as work that examines broader issues of performance prediction and candidate reactions, we outline the development and empirical test of a performance interview that can be used for skill-based promotions. Such interviews allow employees to demonstrate their skills on the job, and thus may be more accepted by employees. We first discuss seven design principles that form the conceptual foundation for the performance interview, followed by the specific steps practitioners can follow to develop a performance interview. Finally, using a sample of 230 auto parts manufacturing employees, we provide empirical evidence for the high reliability, validity, and positive candidate reactions to the performance interview.

1. Introduction

The pace of technological change in work continues unabated. Although there is a strong need for workers to adapt to technological change in order to remain competitive, there still exists a persistent gap between the skills needed and the skills possessed by workers (e.g., as evidenced by the passage of the National Skills Standards Act in 1994; Torraco, 2007). For example, in its 2005 Skills Gap Report, the National Association of Manufacturers and Deloitte Consulting reported that >80% of the 800 manufacturers surveyed experienced a 'moderate' to 'severe' skill shortages. Given this, over the past 20 years organizations have sought to implement a variety of systems to foster skill development. These programs seek to identify the skills that support an organization's business strategy, and then implement a human resource (HR) system designed to foster or develop these skills among organizational members. Two prominent HR strategies that have emerged are skill-based pay systems and skill-based training programs. These programs have been shown to enhance skill development and other important organizational outcomes (Lawler, 1990; Murray & Gerhart, 1998; Smith-Jentsch, Salas, & Baker, 1996).

Yet there is less guidance in the literature about how to promote employees on the basis of skills when compared with other skill-based HR systems. Criteria commonly used for promotion purposes only indirectly assess the skills needed to perform the next higher job. For example, paper and pencil knowledge tests are commonly used. These directly measure knowledge, but knowledge does not necessarily indicate that someone has actual skill in performing the job, particularly if the knowledge test focuses primarily on declarative knowledge. In addition, employees often dislike taking

paper and pencil tests (Hausknecht, Day, & Thomas, 2004; Schmidt, Greenthal, Hunter, Berner, & Seaton, 1977), particularly if they have demonstrated success on the job or similar jobs in the past.

As another example, many promotion systems utilize the amount of time an individual has spent in the job as an important factor, which has been shown to be modestly related to job performance (McDaniel, Schmidt, & Hunter, 1988; Schmidt & Hunter, 1998). Although longer tenured employees tend to have slightly greater job knowledge (Schmidt, Hunter, & Outerbridge, 1986), tenure is only an indirect indicator of knowledge.

Given the limitations inherent in some of the commonly used promotional criteria, we sought to develop an approach that builds on the extensive literature on employee assessment and explicitly focuses on the skills employees will need in their next assignment. Drawing from research on structured interviewing and performance testing, as well as work that examines broader issues of performance prediction and candidate reactions, we developed an assessment we call a 'performance interview.' We chose this term for two reasons. First, the format of the assessment is a structured interview. Multiple interviewers use standardized questions and rating scales to judge the quality of the promotional candidate's answers.

Second, it is a performance test that is conducted at the actual job site. Broadly defined, performance tests are 'standardized measures of behavior whose primary objective is to assess the ability to do rather than the ability to know' (Cascio & Phillips, 1979, pp. 751–752) and are highly valid predictors of job performance (Hunter & Hunter, 1984). Long studied under the label of 'work sample' (Brugnoli, Campion & Basen, 1979; Campion, 1972), hands-on performance tests (HOPTs) (Carey, 1992, 1994; Carey & Mayberry, 1992) and walk-through performance tests (Hedge & Teachout, 1992; Kraiger & Teachout, 1991) represent two recent innovations that have been explored primarily in military settings. We extend this previous work in performance testing by developing a performance interview that enables a promotional candidate to demonstrate his or her performance capability in the work setting.

The original idea for a performance interview was suggested by an employee committee in the organization we studied. They protested the use of paper-and-pencil tests of job knowledge and instead suggested (in the words of one employee), 'If you want to see if I can perform the job, why don't you just come out on the job with me and I will show you.' The performance interview helped determine whether an individual was ready to move to the next major job level and can be used in most skill-based promotion situations.

We begin our discussion by outlining seven design principles extracted from the assessment literature that were used to develop the performance interview. Following this, we outline the specific steps involved in developing a performance interview that can be used in many assessment contexts. Next, we present evidence for the reliability, validity, and candidate reactions of a performance interview that we developed. Finally, we close with a discussion of other advantages and costs of a performance interview.

2. Design principles

Over the years, much has been written with respect to how to (1) maximize performance prediction and (2) develop an assessment that is accepted and viewed as fair by employees. Because we were explicitly focusing on the assessment of job-related skills, we examined the literature for principles that could guide the development of a skill-based assessment technique that was viewed positively by those completing the assessment. We identified seven such design principles below.

2.1. Design principle #1: use a task-based competency model

Gaining a thorough understanding of a job and its requirements has long been viewed as essential before developing an assessment technique. This is typically accomplished by conducting a job analysis to identify the underlying knowledge, skills, abilities, and other characteristics (KSAOs) needed for successful job performance. It is perhaps the most basic design principle that is universally endorsed, in part because it represents a key link to job relatedness (essential to legal defensibility under the Uniform Guidelines on Employee Selection Procedures, 1978).

Competency modeling is a recent extension of traditional job analysis methodologies (Shippmann, Ash, Battista, Carr, Eyde, Hesketh, Kehoe, Pearlman, Prien, & Sanchez, 2000). Although there are many different definitions of competency modeling, in our conceptualization, competency modeling is a process for modeling the growth in employee responsibilities (i.e., major tasks and duties) and knowledge/skill across job levels. One of the key features of a competency modeling approach is that it explicitly takes into account employee characteristics valued by the organization (thus reflecting a strategic orientation) and can serve to align key HR systems so they reinforce or support one another. Given the need to model the growth across job levels in the promotional system, it is essential to identify the key tasks and knowledge and skill needed to successfully perform the tasks at *each level*. This is important because, 'skills cannot be defined apart from some performance domain involving the

acquisition and application of certain kinds of knowledge' (Mumford, Peterson, & Childs, 1999, p. 50).

Such a focus on a performance domain suggests that one should not immediately shift from the task level to the more general knowledge and skill level when developing an assessment. This is important for three reasons. First, the tasks of the job define the performance domain in terms of the behaviors exhibited on the job. It is often the case that when conducting a job analysis for assessment purposes the primary objective is to, 'jump very quickly to a generalized statement of skills and abilities rather than remaining on the behavioral level' (Wernimont & Campbell, 1968, p. 374). Second, identification of job tasks is key to establishing the content validity of an assessment, which cannot only aid legal defensibility, but also reactions of employees because the assessment appears like the job. Third, a focus on tasks enables one to explicitly define the work that is performed at each of the levels in the promotion system. This is crucial when developing a skill-based promotion assessment because it allows one to measure the capability of incumbents to perform the key activities that reside at the next highest job level.

2.2. Design principle #2: adopt a behavioral consistency approach

Wernimont and Campbell (1968) distinguished between the use of tests as signs (or indicators) of predispositions to behave in particular ways as compared with samples of behavior in predicting job performance. They suggested that the use of samples would serve as better predictors of future performance and advocated what they termed a behavioral consistency approach. The goal of such an approach is to create a, 'point-to-point correspondence between predictor and criterion' (Schmitt & Ostroff, 1986, p. 91). There are at least two advantages of the behavioral consistency approach. First, it entails a consideration of the dimensions of the actual behaviors exhibited on the job. This creates a tight linkage between job performance constructs and the operational measures used. Second, it helps ensure content validity of an assessment by making it more likely that key job tasks are included in the assessment. A behavioral consistency approach would be useful in a skill-based promotion assessment context because it allows employees the opportunity to demonstrate their skill on the job-related behaviors and tasks that are performed at the higher level job.

2.3. Design principle #3: use maximum performance measures

Although assessments can be classified in a number of different ways, one common distinction is between assessments that attempt to measure the maximum performance of an employee vs assessments that attempt to measure what an employee will typically do in a given situation (Cronbach, 1970). Generally speaking, maximum performance measures reside in the ability or knowledge domain, whereas typical performance measures reside in the interest or personality domain (Sackett, Zedeck, & Fogli, 1988). There is compelling evidence that maximum performance measures are much better predictors of job performance than typical performance measures (Schmidt & Hunter, 1998; see also the special issue on typical and maximum performance edited by Klehe, Anderson, & Viswesvaran, 2007). The use of maximum performance measures in skill assessment is common because it enables an understanding of the depth of knowledge and capability in a particular domain. This is important because one wishes to promote the most capable person. In addition, there might be certain minimum standards that need to be met before promotion. If candidates cannot perform at a predefined level of mastery, then they should not be promoted.

2.4. Design principle #4: measure declarative and procedural knowledge

Distinctions have commonly been made between declarative and procedural knowledge (McCloy, Campbell, & Cudeck, 1994). Declarative knowledge reflects knowledge of facts, rules, principles, and procedures. Procedural knowledge is the capability attained when knowing what to do (declarative knowledge) 'has been successfully combined with knowing how' (McCloy *et al*., 1994, p. 494). To the extent that an assessment can measure both forms of knowledge, it will more adequately capture the full capability of an individual. This is important in skill-based assessments because the underlying knowledge of both what to do and how to do it is necessary for the demonstration of a particular skill. Explicitly including both declarative and procedural knowledge elements is also helpful because assessment failures could be due to weaknesses in either type of knowledge. Resulting candidate feedback can more specifically target the knowledge deficit.

2.5. Design principle #5: ensure psychological and physical fidelity

In discussing the process of establishing the content validity of a selection or promotion tool, Goldstein, Zedeck, and Schneider (1993) make a distinction between physical and psychological fidelity. An assessment can be said to have high-physical fidelity to the extent it represents the actual tasks performed on a job. The psychological fidelity of an assessment, on the other hand, is dependent upon the extent to which the

knowledge, skills, and abilities needed to perform job tasks are elicited by the assessment. Many assessments have low physical fidelity and therefore rely only on psychological fidelity to ensure content validity. Assessments that can have both psychological and physical fidelity are more likely to have high levels of content validity.

2.6. Design principle #6: consider procedural justice rules

The assessment literature has historically been concerned with the reliability and validity of assessment tools. It is only recently that attention has been given to the importance of the perceived fairness of assessment processes (Gilliland, 1993). Such a focus on fairness does not minimize the importance of reliability and validity, but simply adds another aspect of an assessment process that is important to consider, in part because perceived fairness has been shown to be related to important organizational outcomes such as applicant attraction, intent to accept a job offer, and referring other candidates (Bauer, Truxillo, Sanchez, Craig, Ferrara, & Campion, 2001; Maertz, Bauer, Mosley, Posthuma, & Campion, 2004; Truxillo, Bauer, Campion, & Paronto, 2002). Gilliland (1993) articulated 10 procedural justice rules that would be likely to improve the perceived fairness of a selection system, including formal characteristics (job relatedness, opportunity to perform, reconsideration opportunity, and consistency of administration), explanation (feedback, selection information, and honesty), and interpersonal treatment (interpersonal effectiveness of administrator, two-way communication, and propriety of questions). If a promotional assessment can be explicitly designed with these rules in mind, employees are much more likely to accept the assessment and the decisions that result. Fairness perceptions in promotional contexts are particularly critical. In hiring contexts, candidates who fail the assessment process do not join the organization, whereas promotional candidates who fail the assessment process remain in the organization. As a consequence, their judgements of the fairness of the assessment process may have a larger impact on the organization because they remain employees.

2.7. Design principle #7: use multiple components of interview structure

An employment interview is perhaps the most commonly used assessment technique, with structured interviews being particularly effective. There are a number of advantages associated with interviews that would recommend their use for skill-based assessments. First, structured interviews have demonstrated excellent reliability and validity (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994). Second, interviewees generally have positive reactions to interviews when compared with other assessment techniques (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). This is particularly important in promotional contexts because of the fact that some employees will not be promoted, and when outcomes are unfavorable, procedural justice concerns become much more important (Brockner & Wiesenfeld, 1996). Third, the interview is a highly flexible assessment, in which many different characteristics can be assessed, including job-related skills (Huffcutt, Roth, Conway, & Stone, 2001). Finally, interviews allow employees to provide answers in their own words and actions. This has the dual advantage of producing positive interviewee reactions and being a highly flexible assessment tool that can be applied to a range of jobs across a wide variety of settings.

Campion, Palmer, and Campion (1997) have clarified the different ways an interview can be structured by developing a model of 15 components of interview structure. This includes structuring both the interview content and the evaluation process. Ways to structure interview content include using a job analysis, asking the same questions of all candidates, limiting prompting, using higher quality questions, conducting a longer interview, controlling ancillary information, and not allowing candidates to ask questions during the interview. Ways to structure the evaluation process include rating each answer separately or using multiple scales, using anchored rating scales, taking detailed notes, using multiple interviewers, using the same interviewers across candidates, not allowing discussion of candidates between interviews, providing interviewer training, and utilizing statistical prediction. In the performance interview we developed, we enhanced the structure by using many of these components, although it is important to recognize that it is not always necessary (or desirable) to include all these components of structure. The choice of level of structure should be guided by the particular application.

3. Organizational setting

Before discussing the development and administration of the performance interview, it is helpful to describe the setting within which the performance interview was developed. The organization is located in the midwest United States and supplies engine components to a multinational automaker. Eight departments manufactured and assembled a variety of automobile engine components. All production employees were referred to as 'associates,' and there were few external status differentiators. Within each department, there was one

overall coordinator, several team leaders (TLs), and several technical associates (TAs) in addition to the production associates. The performance interview was designed to encourage personal and career growth and increase associate skill level. Associates were allowed to review the performance interview materials before being assessed. Such a policy enabled associates to better prepare and thereby develop their skills in anticipation of completing the performance interview.

A committee of associates developed the program during weekly meetings over a 10-month period. The committee served many roles, including helping conduct the job analysis and developing the competency model, discussing design issues and alternatives, communicating to the other associates about the program and bringing their feedback to the committee, solving problems, and agreeing on the major decisions about the program. The performance interview was used to assess whether an associate had enough job-related skill to be promoted (individuals passing the performance interview received a promotion in title as well as a promotion in pay). Performance interviews were customized for each of the eight departments.

4. How to develop and administer a performance interview

The seven design principles formed the foundation for the development of the performance interview. In this section, we first discuss how the performance interview was developed and then describe how it was administered.

4.1. Competency model

The first step in developing the performance interview was the identification of the specific tasks performed and the KSAOs needed to effectively perform the jobs. Because these were production associate jobs, we started the job analysis process by surveying 17 senior production associates in order to develop a 'starter list' of important responsibilities and KSAOs. The senior production associates generated and then rated the importance and time spent of a number of job responsibilities. They also rated the level of skills, abilities, and work styles needed to perform the job.

The associate committee then used the job analysis data to develop competency models that consisted of a matrix of how responsibilities and skills changed by job level, from associate, to senior associate, to expert levels (see Tables 1 and 2 for illustrations). These matrices were developed by first identifying and defining the associate and senior associate levels (both levels already existed). This served as the starting point for defining what an expert level associate might do (because there was no existing expert level job). In facilitating group brainstorming sessions with the associate committee, we focused on a single responsibility or skill at a time, developed a definition of the responsibility or skill at each job level, and then committee members vetted the draft matrices with other associates in their respective departments.

Responsibility and skill matrices were developed for each of the eight departments. This tailored the matrices to the work in a given department very closely. Each matrix included between 9 and 13 major responsibilities and about 18 major skill or knowledge areas (Table 3). These matrices were reviewed and revised by virtually every production associate in the company through discussions and meetings with the members of the associate committee. This broad involvement served to enhance commitment to the program. The responsibility by level matrices detailed a series of tasks for each of the major responsibilities across each of three job levels in each department. The skills matrices combined the KSAOs into categories and then defined the requirements for each level in each department that were needed upon entry to the job. Thus, the resulting responsibility and skill matrices (i.e., the job analysis output) were tailored to each level.

Both sets of matrices were used to develop the performance interview as well as the job performance measure used to validate the performance interview. This ensures that the test items directly assess an associate's ability to perform actual job-related production tasks. For example, the responsibility matrices directly determined the interview questions to be asked and the skilled behaviors that needed to be observed, and the skill matrices informed the rating scales. This also ensured that the performance interview and job performance measure were aligned in that both measured the same constructs and defined performance at each job level in a similar way.

4.2. Performance interview content

The performance interview was developed with the input of a wide range of subject matter experts (SMEs) that included production associates, TLs, TAs, and coordinators in addition to the associate committee. In gathering preliminary information from the SMEs, the question content and types were developed. The goal was to gather information on what knowledge is needed to perform the job (i.e., declarative knowledge) as well as how to perform the job (i.e., procedural knowledge and skill). Based on the job analysis, information was gathered from SMEs on the following major content areas: (1) Performing procedures; (2) Troubleshooting; (3) Maintenance; (4) Quality (e.g., checks

Table 1. Sample responsibility by level matrix

Responsible for:	Associate-level	Senior associate-level	Expert associate-level
Machine set-up and model changes	• Assists and learns machine set-up, i.e. start-up, shut down, die changes • Prepares and warms up machines correctly • Performs machine checks • Understands die change schedule • Assists with die changes	• Performs machine set-up • Performs model changes with minimal supervision • Makes adjustments • Changes major components with minimal supervision, i.e. changes blades, jigs, tools, etc.	• Checks calibrations on machines • Changes major components without supervision • Sets up without supervisor need for follow-up • Researches and troubleshoots machine set-up
Operating machines	• Knows and performs basic operations to keep machine(s) running • Loads and unloads machines correctly • Meets production quota at specific machine	• Monitors and operates several machines at one time (depending on area) • Operates different types of machines at different times • Makes minor machine adjustments	• Troubleshoots machine operation to reduce errors, increase productivity, increase efficiency, improve quality, etc. • Run all machines in area (at different times) • Consistently operates machine to its maximum capacity
Machine recovery	• Analyzes and determines correct recovery steps • Resets machines to start position	• Makes minor adjustments • Takes initiative to learn and assist in complex machine recoveries	• Makes major adjustments • Performs complex machine recoveries • Troubleshoots why machine went down • Suggests ways to eliminate need for machine recovery
Preventative maintenance	• Performs basic machine cleaning • Checks oils, fluids, coolants, paper, belts, etc. • Learns PM • Performs COP procedures	• Changes oils, filters, brushes, springs, papers, belts, etc. • Inspects machine condition • Performs PM procedures • Fills out TPM tags as needed	• Conducts or assists complex PM procedures • Verifies that proper PM has been performed • Troubleshoots machine problems and improvements in PM procedures • Knows how and why to maintain PM schedule

Note: This is only a partial matrix for one department.

Table 2. Sample skill by level matrix

Skill/Knowledge	Associate	Senior associate	Expert associate
Production process and operation knowledge	• Knowledge of operation standards • Knowledge of start-up procedures • Knowledge of production goals and requirements • Knowledge of supplies needed (e.g., perishable parts, clothing, etc.) • Knowledge of machine shutdown procedures	• Knowledge of machine recovery procedures • Knowledge of finished products (e.g., assembly procedure, operation, etc.) • (Basic) Knowledge of different model types • (Basic) Knowledge of production processes, flows, systems, layout, etc., both in own area and in entire plant	• (Advanced) Knowledge of different model types • (Advanced) Knowledge of production processes, flows, systems, layout, etc., both in own area and in entire plant
Problem-solving skills	• Problem identification	• Information gathering • (Basic) Identification of key causes	• Testing countermeasures • (Advanced) Identification of key causes
Technology knowledge	• Knowledge of machining technology • Knowledge of technical terms • Knowledge of hydraulics, pneumatics, etc.	• Knowledge of robotics • Knowledge of welding processes and techniques • (Basic) Knowledge of preventative maintenance (e.g., machines, tools, power supplies, electrical circuits, etc.)	• (Advanced) Knowledge of preventative maintenance (e.g., machines, tools, power supplies, electrical circuits, etc.) • Knowledge of schematics (e.g., electrical drawings, parts drawings, etc.)

Note: This is only a partial matrix for one department.

Table 3. Major responsibilities and knowledge, skills, and other characteristics

Responsibilities	Knowledge, skills, and other characteristics
1. 5S (cleaning)	1. Communication skills
2. Checking quality	2. Dependability
3. Inventory control	3. Energy
4. Machine maintenance	4. General education/learning skills
5. Machine recovery	5. Information management and resources knowledge
6. Machine set-up	6. Initiative
7. Manually assembling parts	7. Organizational knowledge
8. Manually process parts	8. People skills
9. Operating machines	9. Physical abilities
10. Paperwork	10. Problem-solving skills
11. Preventative maintenance	11. Production process and operation knowledge
12. Quality control and paperwork	12. Production process skills
13. Safety	13. Quality specifications, equipment, and procedures knowledge
14. Self-development	14. Reasoning/thinking abilities
15. Training others	15. Safety knowledge
	16. Stress tolerance
	17. Teamwork
	18. Technology knowledge

conducted on the machine, checks conducted by the associate, part specifications, etc.); (5) Identifying machine parts and functions; (6) Identifying product parts and functions; (7) Identifying tools (including quality tools); (8) Reading drawings and schematics; and (9) Recording information in reports.

Information was also collected from SMEs to develop the following types of questions: (1) Demonstrating procedures; (2) Describing procedures; (3) Describing consequences (e.g., 'What happens if . . .?'); (4) Describing and demonstrating quality checks; (5) Reading a drawing or report; and (6) Identification of items (e.g., tools, parts, reports, etc.). Questions were then written to tap into these important content areas using these types of questions.

When gathering information from SMEs, the following plan was used (the imperative comments are directed toward the individual collecting the information to develop the performance interview):

1. Start in one area of the plant at a time.
2. Review the types of questions that can be asked.
3. Review the responsibility and skills matrices with SMEs. Pay attention to the distinctions between the

levels (e.g., Associate, Senior Associate, and Expert Associate).
4. Develop questions based on interviewing job incumbents and other SMEs at the job site.
5. Develop questions for each responsibility one at a time.
6. When interviewing the SMEs, ask for information that will lead to interview questions.
7. Give the SME an example or two of the kinds of questions needed.
8. Start with the Senior Associate questions, then do the Expert Associate questions, but collect some information on both levels in each interview (Note: We did not ask for associate-level questions because promotions were from associate to senior associate and senior associate to expert levels).
9. For Associate vs Senior Associate, possibly ask questions such as:
    - 'How could you tell if a person knows how to perform the job?'
    - 'What do they know?'
    - 'What do they know how to do?'
10. For developing Senior Associate vs Expert Associate test items, possibly ask:
    - 'How could you tell if a person was a real expert on the job?'
11. Keep in mind that the differences between levels may be as much the depth of the question, as opposed to different types of questions.
12. Strive to collect at least several questions per interview with each SME.
13. Especially focus on collecting information from SMEs in supervisory roles (e.g., TLs, TAs, coordinators, and managers).
14. Take extensive notes roughing out the questions, and then refine the questions later.
15. It may not be possible to collect every possible question. Instead focus on collecting a good sampling of questions across the different responsibilities.

Once draft performance interviews were developed, they were pilot tested with every coordinator, TL, and TA (approximately 10 per department). The pilot test involved an introduction to the performance interview, a review of the performance interview procedures, and an abbreviated performance interview given to select associates. The pilot test resulted in many substantive changes to the performance interview procedures and materials. It also served to train all leadership personnel who would be administering the performance interview.

The development process resulted in a performance interview for each department. Each performance interview contained about 12 'items' associated with each of 9–13 major responsibilities, covering all major areas in a department (about 120–150 questions). An example of an expert-level performance interview form (for a single responsibility) is provided in Figure 1. As shown, a number of specific questions can be asked. These questions (along with the general follow-up questions) are designed to gain an understanding of the depth of an associate's skill and knowledge by having the associate demonstrate or explain what he or she would do at the job site. In this way the performance interview consisted of arrays or patterns of questions from which an interviewer could select. The questions helped to directly determine if the associate could perform the tasks at the next higher level. The tasks associated with a particular responsibility (directly from the responsibility matrix) were also included to remind the panel of the job activities the associate is expected to be able to perform and that they should observe in the interview. It also served as an explicit link to the competency model, thus ensuring alignment. A place for notes was provided to facilitate recall of an associate's answers in post-interview discussions. Finally, an anchored rating scale was developed to provide raters with an appropriate frame of reference for their ratings.

4.3. Performance interview administration

The performance interviews were administered by a panel of three SMEs. The panel typically included one TL, one TA, and a trainer or another expert-level associate (depending on availability). The performance interviews took between 2 and 3 h to complete, including a 15-min break in the middle. The entire process (including both pre- and post-meetings) typically took a total of 1/2 day for the performance interview panel.

4.3.1. Preplanning meeting

Before administering the performance interview, the panel held a preplanning meeting to discuss the job responsibilities, the tasks, and the questions. The panel was expected to ask most or all of the questions. The panel was also expected to ask follow-up questions that challenged the associate and probed the associate's entire range and depth of skill. In particular, the panel was instructed to ask some questions from areas where the candidate has special expertise and some from areas where the candidate may be 'weaker' (i.e., have less knowledge and skill). In addition, the panel developed a plan for the areas on which to test the associate, including which machine(s), process(es), or line(s) to be included in the interview.

Responsibility to be Assessed: Machine Set-up

Test Associate on the Following Workstations/Machines/Lines: _____

(must be able to perform all of these tasks)

Basic Associate-Level Tasks	Additional Senior Associate-Level Tasks	Expert-Level Tasks
• Assists/learns machine set up (with supervision) • Prepares and warms up machines correctly • Performs machine checks	• Assists/performs model changes (as appropriate for position) • Insert pins • Assists machine adjustments and changes • Assists tool changes (as appropriate for position)	• Assists/performs tool changes depending on line • Performs model changes • Performs machine adjustments

Initial Questions & Follow-up Questions

Ask these questions for this responsibility, focusing on all tasks listed above; ask follow-up questions to fully assess the depth of the associate's knowledge and skill; take 5-10 minutes *for each responsibility* ; skip questions that do not apply

- How do you set up this machine?
- What machine checks do you need to do?
- How do you insert the pins?
- What are some special considerations in setting up this machine?
- What tool changes are made to this machine?
- How are tool changes made on this machine?
- What adjustments are made to this machine?
- How would you make adjustments to this machine?
- What adjustments should you not make to this machine?
- How do you perform model changes on this machine?

General Follow-up Questions

- What else would you do or check?
- What is the next step?
- Can you explain why?
- What would you do if that did not work?
- What could have caused the problem?
- What are you thinking about/looking for as you run this piece of equipment?
- Do you have any other ideas?
- Do you have anything else to add?

Notes & Comments about Associate's Answers (important for justifying ratings): _____

_____

_____

_____

Associate's Overall Skill and Knowledge on this Responsibility (Circle Your Rating)

[½ point ratings (2.5, for example) can also be given]

1 ——————|—————— **2** ——————|—————— **3** ——————|—————— **4** ——————|—————— **5**

1	3	5
• Incorrect procedure	• Some mistakes in procedure	• Fully correct procedure
• Low troubleshooting skill	• Moderate troubleshooting skill	• Very high troubleshooting skill
• Low technical knowledge	• Moderate technical knowledge	• Extensive technical knowledge
• Slow & inefficient	• Moderate speed and efficiency	• Quick and efficient
• Low quality/attention to detail	• Moderate quality/attention to detail	• High quality/attention to detail
• Does not mention safety	• Briefly/indirectly considers safety	• Explicitly considers safety

Figure 1. Sample performance interview form.

4.3.2. Conducting the performance interview

The performance interview was conducted in the actual work area at the equipment that is relevant for a particular responsibility. The performance interview was conducted at times when it was convenient for the interviewers and associate, as well as when it would

not be disruptive to the plant operations (e.g., during maintenance or other down-time).

Associates were asked questions on one responsibility at a time. For example, if an associate was being tested on the responsibility of machine setup, the test questions and answers focused on that responsibility only, and not other responsibilities (e.g., operating machines, machine recovery). To begin the interview, panelists went to one of the associate's areas of special expertise first (e.g., machine, line, or responsibility), then to one of the 'weaker' ones, and finally to one in between. After that, they started over in the same order. Because the performance interview was designed to assess the extent to which an associate can perform activities related to the next higher level, the questions asked during the performance interview involved activities performed in the next higher-level job.

Questions on one responsibility were normally answered (including all follow-ups) in about 8 min. In other words, multiple specific questions were asked for each responsibility (there were 12 responsibilities), but all these questions and answers took approximately 8 min to answer for each responsibility. The panel sometimes found it was necessary to expand the topic slightly when asking follow-up questions in order to probe the depths of the associate's knowledge and skill. Follow-up questions were used: (1) to narrow the focus of the question if the associate was giving only general answers; (2) to explore potential weaknesses that were identified by the associate's previous answers; (3) to see if an associate recognized an incorrect fact or question; and (4) to find out how much the associate knew. The goal of follow-up questioning was to probe the depth of an associate's knowledge. It is essential to ask enough follow-up questions to fully assess the associate's skill level. Normally, follow-up questions were asked until all the panel members felt they had adequate information to make a rating on the responsibility. This was often at the point when no one else had any remaining questions to ask.

Depending on the responsibility and specific question, we found that it was often most appropriate for an associate to actually perform the task to demonstrate his or her knowledge. For other responsibilities and questions, it was most appropriate for the associate to simply explain the answer and refer to the relevant aspects of the task. The decision on whether to have the associate perform the task or explain the answer was based on a consideration of the responsibility being assessed, the extent to which it was easier to explain or demonstrate the underlying knowledge or skill, and the feasibility (given ongoing production needs) of actual performance. Finally, it was sometimes appropriate to ask the associate to explain how to perform the task as if training someone else.

With multiple people asking questions and the associate providing multiple answers, we found that it was sometimes helpful to occasionally stop and review. This was often needed because of the complexity of the answers, and it helped keep the associate and the panel organized and on track. At these times, it was helpful for the panel to briefly summarize what the associate had said up to that point. This often stimulated further follow-up questions and helped the panel understand the associate's response.

Panel members were encouraged to take notes on the answers. Extensive notes were not required, but they needed to be sufficiently detailed to be able to reconstruct the answers afterwards in case of a disagreement on the rating. We found it useful to take notes on key points the associate answered correctly, points the associate answered incorrectly, and points the associate failed to address. Circling the points of an excellent answer on the answer sheet to indicate that the candidate made them was also permissible.

Finally, when completing the performance interview, associates were given access to any resource material that would normally be available on the job (e.g., technical manuals). However, they were not allowed to ask other associates or coworkers for help.

4.3.3. Scoring

A key, but often neglected, aspect of content validity concerns the scoring of assessment performance. As suggested by Guion (1978, p. 501), an assessment is content valid to the extent that it both adequately samples the content of the job *and* performance on the assessment yields scores that, 'validly reflect the underlying skill, ability, or quality of performance' needed in the job itself. This was achieved in the present study by explicitly linking the performance interview ratings to the competency model.

One rating per responsibility was made, which took into consideration the main questions and all the follow-up questions. Ratings were made independently by each of the three panel members. We asked panel members to make their rating (or a tentative rating) at the time the associate finished answering the questions on a responsibility, but before the panel moved to the next responsibility. High-scoring answers required considerable depth, breadth, and comprehensiveness. As such, we reminded the associate at the beginning of the performance interview to be sure to consider all aspects of the job when answering the questions. When the panel had to remind the associate of many of the key aspects of a given machine or line, the ratings were lower.

The panel met after all the questions and ratings were complete. During the meeting, differences in ratings were resolved. Small differences, defined as one point or less, were simply averaged. Larger differences, defined as 1.5 points or larger, were discussed to consensus or to within a one-point difference. Total

performance interview scores were computed by averaging across all interviewers and all responsibilities. Associates were then informed as to whether they had passed the interview.[1] For those who did not pass, a detailed explanation of the main areas of weakness was provided by the performance interview panel. In this organization, any associate who felt he or she possessed the next higher level skills was eligible to take the performance interview and therefore could be promoted to the next higher level.

5. Evidence of reliability, validity, and procedural justice perceptions

Although the development process yielded a highly content valid assessment, criterion-related validity evidence was also collected.

5.1. Sample and procedure

As mentioned, the performance interview was developed at a midwestern auto parts manufacturer. The validation sample consisted of 230 current associates. Each associate completed the performance interview, and a multisource feedback measure of job performance was completed by three supervisors and three peers. Associates also self-rated their job performance. Fifty-three percent of the sample were female, and the average tenure was 8.95 years. There were 10 minority members. A randomly selected subsample of 30 associates involved in the validation study was asked to evaluate the procedural justice of the performance interview, the multisource feedback measure of job performance, and a written job knowledge test.

5.2. Measures

5.2.1. Performance interview

As described above, the performance interview consisted of 12 responsibility ratings by three independent panelists based on a 3-h interview involving over 100 specific questions. (Note that each individual question was not rated. Rather, ratings were made on the responsibility as a whole based on the answers to specific interview questions.) Internal consistency reliability was .86. The formulas provided by Bliese (2000) were used to calculate the reliability of a single rater [ICC(1)] and the mean reliability [ICC(2)]. For the performance interview and the job performance ratings, the overall score for that measure (i.e., the overall performance interview score across the dimensions) was used as the dependent measure and the candidate ID as the independent measure in a one-way analysis of variance (ANOVA) framework. This partitions the variance into between and within candidate components. Interrater reliability of the mean of the three raters [ICC(2)] was .87, which was statically significant ($F(229, 688) = 7.97$, $p < .01$). Interrater reliability of a single rater [ICC(1)] was .70.

5.2.2. Job performance

An associate's level of job performance was assessed by gathering evaluations from three of the associate's supervisors and three of the associate's peers on the level of performance across the key dimensions identified in the skill and responsibility matrices (ranging from 13 to 19 items). In order to avoid predictor–criterion contamination, no members of the performance interview panel provided job performance ratings on the candidate. The job performance measure assessed the level at which the associate was currently performing (i.e., associate, senior associate, or expert) in terms of the tasks contained in the skill and responsibility matrices (this job performance assessment was independent of the actual level a person was currently assigned). This was accomplished by developing anchored rating scales that reflected potential performance at all possible job levels (i.e., associate, senior associate, and expert). Because of this, the job performance measure indexes an associate's current performance in terms of whether it is at a level consistent with the associate's current level, or a level lower or higher than the level an individual was formally assigned. Therefore, the job performance ratings enabled an independent assessment of whether the associate was performing at the level to which the promotion was being sought.

Associates also rated their own job performance on the same scales. Internal consistency reliability was .84. Interrater reliability of the mean of the six raters (without self-ratings) was .80, which was statically significant ($F(229, 1379) = 5.00$, $p < .01$). Interrater reliability of a single rater was .40.

5.2.3. Procedural justice perceptions

The perceived fairness of the performance interview, job performance survey, and written job knowledge tests (which could have been used in place of the performance interview) was assessed by asking associates to indicate the extent to which the content of the various assessments (on a five-point 'agree' scale) were related to the job on five survey items (e.g., 'This test/survey is related to the requirements of the job;' 'A high score on this test/survey means a person can do the job well'). These items were adapted from Bauer *et al.* (2001) and primarily reflect job-relatedness concerns. Internal consistency of these ratings ranged from .88 to .76.

Table 4. Intercorrelations among study variables

	M	*SD*	1	2	3	4	5
Performance interview	4.10	.47	–				
Other-rated job performance	2.53	.32	.47**	–			
Self-rated job performance	2.53	.68	.18*	.42**	–		
Tenure	8.95	2.11	.02	.05	.02	–	
Gender[a]	–	–	.01	.09	−.07	.06	–
Minority status[b]	–	–	.07	−.16*	−.06	.07	−.10

*$p < .05$. **$p < .01$. [a]Male, 1; Female, 2. [b]Non-minority, 1; Minority, 2.

5.3. Results

As shown in Table 4, the performance interview was significantly related to job performance ($r = .47$, $p < .01$; all correlations are uncorrected). In addition, the performance interview was significantly related to self-ratings of job performance, but at a much lower level ($r = .18$, $p < .05$). This empirical evidence suggests that the performance interview is assessing important job-related skills and knowledge needed to be promoted to the next higher level. In addition, because the individuals who provided performance interview ratings were different from the individuals providing the job performance ratings, we can be more confident that these significant results are not influenced by predictor–criterion contamination. To examine whether differences across departments influenced the results, we conducted regression analyses controlling for department (via a series of dummy codes). We found that controlling for department had no effect on the validity of the performance interview.

Additional analyses were conducted to determine whether there were any subgroup differences. There were no significant differences between male and female associates on the performance interview (Male: $M = 4.10$, $SD = .47$; Female: $M = 4.10$, $SD = .47$) or the job performance ratings (Male: $M = 2.50$, $SD = .33$; Female: $M = 2.56$, $SD = .32$). In addition, we tested for the homogeneity of intercepts and slopes among female and male subgroups by regressing job performance ratings on performance interview ratings. The regression model consisted of performance interview ratings, gender, and the interaction between performance interview ratings and gender. We found no significant differences in intercept (t_{gender} $(228) = 1.38$, NS) or slope ($t_{\text{gender} \times \text{interview}}$ $(228) = -.23$, NS). Thus, we again found no evidence for gender bias.[2]

There were also no significant racial differences on the performance interview (Majority: $M = 4.09$, $SD = .47$; Minority: $M = 4.24$, $SD = .32$), although there were significant (but small) differences in job performance ratings (Majority: $M = 2.54$, $SD = .32$; Minority: $M = 2.29$, $SD = .37$). These analyses by race should be interpreted with caution, however, due to the small number of minority group members in the sample ($N = 10$) and the fact that this is a potentially heterogeneous group of minorities (i.e., this group could include African-Americans, Asians, or Hispanics). Finally, job tenure was unrelated to both the performance interview or job performance. This suggests that learning on the job depends more on ability and motivation than tenure.

In terms of procedural fairness, on a five-point scale (where a '5' indicated the test was considered to be more job-related), the performance interview had an average rating of 3.20 ($SD = .97$), the job performance measure had an average rating of 2.67 ($SD = .79$), and the written job knowledge test had an average rating of 2.92 ($SD = .77$). A one-way ANOVA revealed that the performance interview was judged to be significantly more job-related than either the job performance measure ($d = .60$) or the written job knowledge test ($d = .32$) ($F(1, 76) = 4.13$, $p < .05$). This suggests that associates felt the performance interview, which was the primary determiner of promotions, was the most job related. It is somewhat surprising, however, that the procedural fairness ratings of the written job knowledge test were rated higher than the job performance measure. We are not entirely sure why this was the case, although in this organization there was some skepticism about using job performance ratings for promotional purposes. This skepticism might have been due, in part, to the belief by some associates that the job performance ratings could be subject to manipulation (i.e., associates could be given inappropriately high or low ratings by peers).[3]

6. Summary and conclusion

We have detailed the development and validation of a performance interview, which combines elements of work sample tests and structured interviews, resulting in a highly content valid promotional assessment. It differs from work samples in that candidates do not necessarily perform the task under highly controlled conditions. This provides for a level of flexibility that can be useful in many promotion contexts, in part

because it offers an opportunity to measure proficiency on tasks that are not feasible to measure in a work sample test or not present in a current job. The performance interview differs from structured interviews in that it goes beyond simple descriptions of what a candidate might have done in the past or might do in future situations and instead focuses on hands-on demonstrations. This offers a level of realism and face validity that can enhance perceptions of procedural justice. Also, the performance interview allows the incumbent not only to demonstrate performance on a task but also to describe in detail how he or she would perform it and describe the underlying reasons certain task-related choices are made. As such, it enables a more direct assessment of declarative and procedural knowledge and skill.

We found that the performance interview evidenced good criterion-related validity, positive procedural justice perceptions, and no mean differences between men and women or majority and minority group members (although these analyses by race should be interpreted with caution due to the small number of minority group members in the sample). In addition, we have provided a detailed a step-by-step description of how to develop a performance interview, including the range of choices that can be made as well as sample materials to enable practitioners to develop a performance interview in their own particular organizational context.

We thus extend existing work on HOPT and walk-through performance tests by integrating them with elements of structured interviews. In addition, we add to the limited empirical research on these types of assessments by studying them in an industrial setting. Most of the past research has been conducted in military settings. There are at least three key differences between 'walk-through' or 'hands-on' assessments and the performance interview. First, both walk-through and hands-on assessments are explicitly focused on these assessments as criteria to be used in selection research, whereas the current research focuses on the use of a performance interview as a predictor in promotion contexts. For example, Hedge and Teachout (1992, p. 453) indicate that walk-through performance testing is designed to 'integrate an interview and a hands-on approach to criterion measurement within a work sample framework.' Similarly, Kirk and Brown (2003) treated the walk-through performance test used in their study as criterion reflecting maximal performance. The HOPT literature has a similar focus. For example, Carey (1992, p. 103) notes that '. . . a hands-on performance test (HOPT) has been advocated as the best measure to use as a validation criterion.' Our focus is in the use of the performance interview as a predictor that can be used to make promotion decisions. We feel that such a shift in focus is important because it makes it clear that such assessments can be very helpful in promotion contexts. Perhaps one of the reasons more organizations do not utilize such assessments is because of how they have been described in the past.

Second, the nature of the scoring rubric for walk-through assessments is very different from the scoring rubric for the performance interview. For example, Kraiger and Teachout (1991) indicate that walk-through assessments should use a step-level, dichotomous scoring of responses (in which a person is scored as having correctly or incorrectly completed a particular step in a larger sequence of actions). This entails a very expensive and time-consuming process in determining (1) all the discrete steps involved in a particular job and (2) the correct way to perform each step. As Hedge and Teachout (1992, p. 456) explain:

> Each WTPT task was composed of a series of steps that had been previously weighted on the basis of the importance of that step to the successful completion of the task. These weights were assigned by job experts (senior noncommissioned officers) from each specialty during workshops held prior to data collection. Weights were then summed across all steps for a task, creating a base score for that task. Points for each step scored as correctly performed were summed, divided by the base score, and multiplied by 10. This placed each task score on a 0–10 scale, so that all tasks, regardless of the number of steps, received equal weight in the computation of the total test score.

The performance interview, however, is scored like a traditional structured interview, which is far simpler to create and administer. In addition, it is more flexible in case there are small changes or differences in a particular work area.

Third, the training required for administrators of walk-through assessments is typically very extensive. As detailed by Hedge and Teachout (1992, p. 456), '. . . test administrators received 1–2 weeks of observation and scorer training . . .. Training of the administrators included instruction in observation and evaluation, interviewing, and walk-though administration procedures. Methods of training consisted of lecture and discussion, role playing, and review of videotaped task performance.' Such extensive training is no doubt needed given the complexity of the scoring system. The performance interview, on the other hand, required relatively little interviewer training (about 1 h) yet yielded very high interrater reliability.

In addition to these differences, there are a number of additional positive aspects of the performance interview worth noting. First, by administering the interview at the work site (i.e., machine or line) and allowing candidates the opportunity to use materials typically available when performing the job (e.g., technical manuals, operating instructions), the relationship between interview performance and promotion decision is clear.

Candidates are being assessed on the specific tasks they will need to perform in the higher-level job. This ensures that the measurement of the knowledge and skill needed is at the correct complexity level.

Second, by using SMEs as interviewers, interview training is quite easy. The relatively short interviewer training (about 1 h) produced highly reliable ratings. Third, a key unexpected positive outcome was the competition that developed in the performance interview. Most associates viewed the performance interview as an opportunity to demonstrate the depth of their knowledge and saw it as a challenge. It motivated associates to review the competency models and interview questions before the assessment in order to 'study' for the promotional opportunity (although it should be noted that such study likely restricted the range of responses on the performance interview). As a consequence, the competition aspect had positive implications for the skill development of the workforce. In addition, the performance interview allowed one to determine exactly how much knowledge and skill an employee possessed, which led to useful feedback for those who failed (about 15% failed).

Fourth, because there are many different ways of expressing the same underlying substantive material (as with any 'constructed response' type of testing), the performance interview was a highly flexible assessment tool. As the work or equipment changes, the performance interview can easily adapt. In addition, associates liked the fact that the interviewers were not simply looking for a single right answer. They could answer in their own words or demonstrations. What resulted during the performance interview was an active and dynamic exchange between interviewer and associate.

Fifth, the level of validity found in the current study is considerably higher than previous research. For example, Hedge and Teachout (1992) report correlations between their walk-through assessments and a host of job performance measures, with the majority of correlations in the .20–.30 range. Carey (1992) reported *corrected* correlations between a hands-on assessment and supervisor performance ratings of .27 and .29. Also, Roth, Bobko, and McFarland (2005) conducted a meta-analysis of work sample test validity and found observed mean correlation between work sample tests and measures of job performance of .26 and corrected correlation of .33.

Sixth, for all the reasons mentioned, associates viewed the performance interview as highly job related, fun to do, and a fair way to determine skill-based promotions. This was confirmed by the quantitative data we collected, but also by numerous comments made by associates. Associates viewed it like giving a tour of their work area and job, except in greater depth. Employee acceptance is particularly important in the current non-union manufacturing environment.

Notwithstanding these positive features, there are several potential drawbacks to the performance interview. First, there are likely certain situations in which the performance interview might work better than other situations. For example, the performance interview is likely to work best when employees have the opportunity to learn and develop skills needed for the next level position in the current position. In addition, the performance interview is better suited where major elements of the job lend themselves to an overt performance demonstration. There are some jobs where this is not the case. These are important boundary conditions.

Second, another potential weakness of the performance interview methodology used was that there may be some inconsistency in the administration as each candidate is likely to receive a different set of questions. Although this is a weakness from the perspective of the procedural justice rules (because of compromised consistency), we feel that it is a strength of the assessment process itself because it enables the test to be customized to the strengths and weaknesses of each candidate. Such customization is made possible because the interviewers were knowledgeable about the candidates and their respective strong and weak areas.

Finally, the performance interview is somewhat time consuming to develop and administer. We used an associate committee to develop the performance interview. This committee met once a week for several months, with meetings facilitated by consultants. The materials were iteratively developed over this time by consultants with considerable input from the associate committee. It is difficult to assign an exact cost figure to this development process. To the best of our knowledge, there are no continuing development costs, although we imagine that periodic revision may be necessary if the work equipment or work process changes significantly. The performance interview typically takes 2–3 h to administer, with about 1–2 h of interviewer preparation/discussion time. If three interviewers are used, the performance interview would involve between 11 and 18 h of employee time. In addition, there is a modest amount of administrative time devoted to record keeping. Although this represents a potentially large investment for each promotional assessment, there are several features of the performance interview at the organization studied that serves to minimize the ultimate cost. First, the system in place at the organization is criterion referenced. That is, when associates are ready for promotion, they take the performance interview. If an associate demonstrates mastery, they can be promoted to the higher level. As such, the organization studied does not conduct large numbers of performance interviews that might be conducted in a normative or need-based system. Second, as interviewers become comfortable

with the performance interview, the preparation/discussion time is typically reduced because of familiarity with the process and interview content. Third, given the high interrater reliability of the interviewer ratings, it may be possible to use only two interviewers instead of three. Future research could investigate this and other ways to reduce costs.

The effort and cost required to develop a performance interview, however, should be balanced against the prospects of developing a more skilled workforce. In fact, a key outcome of developing the performance interview in the present organization is that it served to raise the skill level of the work force by (1) defining a higher bar of competence, (2) providing an objective measure that associates accepted, and (3) motivating associates to proactively acquire skills on their own (without investing in a formal training program).

Acknowledgements

Thanks to the attendees of the 2004 Personnel and Human Resource Research Group (PHRRG) meeting in Portland, OR for their comments on an early conceptualization of this research.

Notes

1. The cutoff score for passing the performance interview was determined by the associate committee through a group discussion and consensus process. The passing score represents acceptable performance at the higher job level. In this organization, the passing score was set at 70%.
2. Although it is important to recognize that these are relatively low-power tests (Aguinis, Beaty, Boik, & Pierce, 2005).
3. Another potential limitation of these procedural justice analyses concerns whether any of the 30 individuals who completed the procedural justice measure also failed the performance interview. Unfortunately, we do not know the answer to this question, as we did not control the distribution of this survey. We simply instructed the organization to take a random sample of associates by taking every *n*th associate who had completed the performance interview.

References

Aguinis, H., Beaty, J.C., Boik, R.J. and Pierce, C.A. (2005) Effect Size and Power in Assessing Moderating Effects of Categorical Variables Using Multiple Regression: A 30-year review. *Journal of Applied Psychology*, **90**, 94–107.

Bauer, T.N., Truxillo, D.M., Sanchez, R.J., Craig, J.M., Ferrara, P. and Campion, M.A. (2001) Applicant Reactions to Selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, **54**, 388–420.

Bliese, P.D. (2000) Within-Group Agreement, Non-Independence, and Reliability: Implications for data aggregation and analysis. In: Klein, K.J. and Kozlowski, S.W.J. (eds), *Multilevel Theory, Research, and Methods in Organizations*. San Francisco, CA: Jossey-Bass, pp. 349–381.

Brockner, J. and Wiesenfeld, B.M. (1996) An Integrative Framework for Explaining Reactions to Decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin*, **120**, 189–208.

Brugnoli, G.A., Campion, J.E. and Basen, J.A. (1979) Racial Bias in the Use of Work Samples for Personnel Selection. *Journal of Applied Psychology*, **64**, 19–123.

Campion, J.E. (1972) Work Sampling for Personnel Selection. *Journal of Applied Psychology*, **56**, 40–44.

Campion, M.A., Palmer, D.K. and Campion, J.E. (1997) A Review of Structure in the Selection Interview. *Personnel Psychology*, **50**, 655–702.

Carey, N.B. (1992) Does Choice of a Criterion Matter? *Military Psychology*, **4**, 103–117.

Carey, N.B. (1994) Computer Predictors of Mechanical Job Performance: Marine Corps findings. *Military Psychology*, **6**, 1–30.

Carey, N.B. and Mayberry, P.W. (1992) *Development and Score of Hands-on Performance Tests for Mechanical Maintenance Specialties*. CNA Research Memorandum (91-242). Alexandria, VA: Center for Naval Analysis.

Cascio, W.F. and Phillips, N.F. (1979) Performance Testing: A rose among thorns? *Personnel Psychology*, **32**, 751–766.

Cronbach, L.J. (1970) *Essentials of Psychological Testing* (3rd edn). New York: Harper & Row.

Gilliland, S.W. (1993) The Perceived Fairness of Selection Systems: An organizational justice perspective. *Academy of Management Review*, **18**, 694–734.

Goldstein, I.L., Zedeck, S. and Schneider, B. (1993) An Exploration of the Job Analysis–Content Validity Process. In: Schmitt, N. and Borman, W.C.B. *et al.* (eds), *Personnel Selection in Organizations*. San Francisco: Jossey-Bass, pp. 3–34.

Guion, R.M. (1978) Scoring of Content Domain Samples: The problem of fairness. *Journal of Applied Psychology*, **63**, 499–506.

Hausknecht, J.P., Day, D.V. and Thomas, S.C. (2004) Applicant Reactions to Selection Procedures: An updated model and meta-analysis. *Personnel Psychology*, **57**, 639–683.

Hedge, J.W. and Teachout, M.S. (1992) An Interview Approach to Work Sample Criterion Measurement. *Journal of Applied Psychology*, **77**, 453–461.

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

Huffcutt, A.I., Roth, P.L., Conway, J.M. and Stone, N.J. (2001) Identification and Meta-Analytic Assessment of Psychological Constructs Measured in Employment Interviews. *Journal of Applied Psychology*, **86**, 897–913.

Hunter, J.E. and Hunter, R.F. (1984) Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin*, **96**, 72–98.

Kirk, A.K. and Brown, D.F. (2003) Latent Constructs of Proximal and Distal Motivation Predicting Performance Under Maximum Test Conditions. *Journal of Applied Psychology*, **88**, 40–49.

Klehe, U., Anderson, N. and Viswesvaran, C. (2007) More than Peaks and Valleys: Introduction to the special issue on typical and maximum performance. *Human Performance*, **20**, 173–178.

Kraiger, K. and Teachout, M.S. (1991) *Application of Generalizability Theory to the Air Force Job Performance Measurement Project: A summary of research results*. AFHRL-TR-90-92. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Lawler, E.E. (1990) *Strategic Pay: Aligning organizational strategies and pay systems*. San Francisco: Jossey-Bass.

Maertz, C.P., Bauer, T.N., Mosley, D.C., Posthuma, R.A. and Campion, M.A. (2004) Do Procedural Justice Perceptions in a Selection Testing Context Predict Applicant Attraction and Intention Toward the Organization? *Journal of Applied Social Psychology*, **34**, 125–145.

McCloy, R.A., Campbell, J.P. and Cudeck, R. (1994) A Confirmatory Test of a Model of Performance Determinants. *Journal of Applied Psychology*, **79**, 493–505.

McDaniel, M.A., Schmidt, F.L. and Hunter, J.E. (1988) Job Experience Correlates of Job Performance. *Journal of Applied Psychology*, **73**, 327–330.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. and Maurer, S.D. (1994) The Validity of Employment Interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, **79**, 599–616.

Mumford, M.D., Peterson, N.G. and Childs, R.A. (1999) Basic and Cross-Functional Skills. In: Peterson, N.G., Borman, W.C., Mumford, M.D., Jeannerete, P.R. and Fleishman, E.A. (eds), *An Occupational Information System for the 21st Century: The development of O*NET*. Washington, DC: APA, pp. 49–69.

Murray, B. and Gerhart, B. (1998) An Empirical Analysis of a Skill-Based Pay Program and Plant Performance Outcomes. *Academy of Management Journal*, **41**, 68–78.

National Association of Manufacturers. (2005). *2005 Skills Gap Report – A survey of the American manufacturing workforce*. National Association of Manufacturers Web site: http://www.nam.org/2005skillsgap (accessed) 11 February 2008

Roth, P.L., Bobko, P. and McFarland, L.A. (2005) A Meta-Analysis of Work Sample Test Validity: Updating and integrating some classic literature. *Personnel Psychology*, **58**, 1009–1037.

Sackett, P.R., Zedeck, S. and Fogli, L. (1988) Relations Between Measures of Typical and Maximum Job Performance. *Journal of Applied Psychology*, **73**, 482–486.

Schmidt, F.L., Greenthal, A.L., Hunter, J.E., Berner, J. and Seaton, F. (1977) Job Sample vs. Paper-and-Pencil Trades Technical Tests: Adverse impact and examinee attitudes. *Personnel Psychology*, **30**, 187–197.

Schmidt, F.L., Hunter, J. and Outerbridge, A.N. (1986) Impact of Job Experience and Ability on Job Knowledge, Work Sample Performance, and Supervisory Ratings of Job Performance. *Journal of Applied Psychology*, **71**, 432–439.

Schmidt, F.L. and Hunter, J.E. (1998) The Validity and Utility of Selection Methods in Personnel Psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.

Schmitt, N. and Ostroff, C. (1986) Operationalizing the "Behavioral Consistency" Approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, **39**, 91–108.

Shippmann, J.S., Ash, R.A., Battista, M., Carr, L., Eyde, L.D., Hesketh, B., Kehoe, J., Pearlman, K., Prien, E.P. and Sanchez, J.I. (2000) The Practice of Competency Modeling. *Personnel Psychology*, **53**, 703–740.

Smith-Jentsch, K.A., Salas, E. and Baker, D.P. (1996) Training Team Performance-Related Assertiveness. *Personnel Psychology*, **49**, 909–936.

Smither, J.W., Reilly, R.R., Millsap, R.E., Pearlman, K. and Stoffey, R.W. (1993) Applicant Reactions to Selection Procedures. *Personnel Psychology*, **46**, 49–76.

Torraco, R.J. (2007) Low-Skilled Adults in the United States: A case of human resource underdevelopment. *Human Resource Development Review*, **6**, 343–354.

Truxillo, D.M., Bauer, T.N., Campion, M.A. and Paronto, M.E. (2002) Selection Fairness Information and Applicant Reactions: A longitudinal field study. *Journal of Applied Psychology*, **87**, 1020–1031.

Uniform Guidelines on Employee Selection Procedures. (1978) *Federal Register*, **43**, 38290–38315.

Wernimont, P.F. and Campbell, J.P. (1968) Signs, Samples, and Criteria. *Journal of Applied Psychology*, **52**, 372–376.