

OPERATIONALIZING THE "BEHAVIORAL CONSISTENCY" APPROACH: SELECTION TEST DEVELOPMENT BASED ON A CONTENT-ORIENTED STRATEGY

NEAL SCHMITT AND CHERI OSTROFF
Michigan State University

This paper delineates a systematic procedure for operationalizing the "behavioral consistency" notion, proposed by Wernimont and Campbell (1968), for a Civil Service job. The steps used in developing selection tests from a content-oriented strategy are illustrated, and the transformation of specific job behaviors into tests related to job content but modified so that inexperienced applicants can complete them is demonstrated. Test reliability and judged content validity of the examinations are presented.

The goal of any selection instrument is to adequately predict future job performance. While the investigator desires to produce the best tests possible, he or she is constrained by practical limitations imposed by the organization. A further concern involves the legal implications: tests must conform to the "Uniform Guidelines on Employee Selection Procedures" (1978). Thus, the aim of any selection device is to be valid, practical, and defensible in court.

One approach advocated to meet these goals in the prediction of job behavior is the "behavioral consistency" model proposed by Wernimont and Campbell (1968). These authors argue for a focus on meaningful and realistic samples of job behavior. The goal is the creation of point-to-point correspondence between predictor and criterion. The behavioral consistency model has served as a theoretical base for work sample tests, miniature replicas of on-the-job behavior. Other selection procedures such as the behavioral consistency method (Mento, 1980; Schmidt, Caplan, Bemis, Decuin, Dunn, & Antone, 1979), the situational interview (Latham, Saari, Purcell, & Campion, 1980), and assessment centers are based on the notion that selection procedures that are behaviorally consistent with job or task requirements are desirable. Certainly the criterion-related validity of these

tests has been well demonstrated (e.g., Asher & Sciarrino, 1974; Robertson & Kandola, 1982). However, a new emphasis on content-oriented test construction and job relatedness issues has emerged due to federal regulations and the establishment of the Equal Employment Opportunity Commission (EEOC) (Guion, 1974). Despite this new emphasis, there is little literature illuminating systematic procedures for identifying job content and their application to selection procedures in a business setting (i.e., Mento, 1980; Robinson, 1981). Further, it has been argued that tests sampling job content can only be used to select people who already know how to handle the job and, thus, cannot be used for many entry-level positions (Guion, 1974).

This paper delineates a systematic procedure operationalizing the "behavioral consistency" approach for a Civil Service job with a relatively small number of job incumbents and low annual turnover. Thus, it offers an illustration of the steps used in developing selection tests from a content-oriented strategy. Further, this paper demonstrates the transformation of specific job behaviors into tests which are related to job content but modified so that inexperienced people can successfully complete the tests.

Procedure

The job for which selection procedures were developed was that of an emergency telephone operator, officially termed a police technician—39 individuals worked in this capacity during the research period. Police technicians are required to answer telephone calls from complainants regarding emergency situations which require police, fire, or medical attention. These workers gather the pertinent information from callers so that the appropriate assistance can be sent. They communicate with police units, dispatching them to respond to the calls, and they monitor location and activities of police units at all times. Further, job incumbents check warrants for arrests, stolen vehicle reports, runaway reports, etc. All of the tasks are performed via an interconnected computer system.

This project consisted of three major components: (1) a job analysis to determine the tasks involved in performing the job of police technician, and the knowledge, skills, and abilities necessary to perform those tasks; (2) the development of selection devices that sample the tasks needed for adequate job performance; and (3) evaluation of the psychometric properties of the tests and their content validity.

Job Analysis

The job analysis procedure used was an approach proposed by Levine (1983, chap. 5), which incorporated ideas from several different approaches. The job analysis of police technicians consisted of four steps: (1) development of task statements; (2) development of knowledge, skills, and abilities (KSAs); (3) rating of the tasks and KSAs; and (4) analysis of the ratings to determine critical task dimensions and KSAs. Before undertaking these steps, the job analysts, in an effort to better understand the job, met with supervisory personnel, who briefed them on the job, observed and listened to workers on the job during two different occasions, and reviewed various training materials and procedure manuals.

Task Generation

Task statements represent the activities and tasks required to perform the job. The general form of the task statement was consistent with task statement structure used in functional job analysis (see Fine & Wiley, 1971). Two meetings with "job experts," workers well experienced in the job, were conducted to generate the list of task statements for the technician job. During the first meeting, these four workers were briefed on the purpose of the project and the generation of task statements was explained. The definition of a task and examples of task statements related to this job were provided. Each group member was instructed to prepare a list, on his or her own, of task statements. When all members were finished, the lists were reviewed by the group. Each task statement was discussed, and necessary revisions and additions were made. The investigators then edited and grouped task statements and compiled a first draft of task statements. A second meeting with a different group of three job experts was conducted to review the task statements. Each task statement was discussed by the groups, and necessary revisions and additions were made.

A revised list was submitted for review to supervisory personnel to ascertain its completeness. Task statements were again revised and edited by the investigators, and a final version of 78 task statements, separated into dimensional categories, resulted. Assignment of task statements to major performance dimensions was done independently by the authors with 90% agreement. Disagreements were resolved by eliminating the task or rewriting it to reduce ambiguity.

TABLE 1

Critical KSA Dimensions and Illustrative Statements

Communications Skills

- Ability to speak on telephone in a clearly understandable manner
- Ability to control conversations in order to acquire information needed to respond to emergencies

Emotional Control

- Can withstand pressure of attending to several emergencies at once and of constant monitoring of activities
- Remains calm even in emergency situations when a caller is hysterical or upset
- Controls emotions in emergency to gather and transmit appropriate information
- Handles emergencies consecutively in an orderly, efficient manner

Judgment

- Ability to judge what constitutes an emergency and requires immediate attention
- Knows when to use common sense vs. procedures
- Asks for help when needed
- Asks pertinent questions to quickly assess and direct help in emergency situations

Cooperativeness

- Can work as part of a team
- Can work with various types of people who differ in life style and work habits
- Is willing to help coworkers
- Ability to take criticism and correction of work

Memory

- Ability to retain information on several situations at once
- Ability to pay attention and remember details of situations
- Ability to recall information from short notes about phone calls
- Ability to concentrate and attend to information

Clerical/Technical Skills

- Ability to spell common words and street names correctly from oral communications
 - Knowledge of typewriter keyboard and ability to type into forms with minimum errors
 - Ability to fill out forms appropriately and with minimum errors
-

Knowledge, Skills, and Abilities (KSAs) Generation.

A meeting with four experienced technicians was conducted to generate the KSAs needed to successfully perform the tasks demanded by the job. During this meeting, copies of the task statements were circulated to group members to review. The group was asked to generate the KSAs needed to perform each of the tasks. The job positions and functional categories were discussed one at a time. The investigators reviewed and edited the information provided, resulting in a list of 54 KSAs. A sample of these KSAs is reproduced in Table 1.

Task and KSA Rating

Two questionnaires were constructed to evaluate the importance of the tasks and KSAs to the job. The task rating questionnaire asked workers to evaluate each task on a 7-point scale in terms of (1) the relative time spent performing this task compared to all other tasks; (2) the relative difficulty of doing this task correctly compared to all other tasks; and (3) the criticality of the task as judged by the degree to which incorrect performance results in negative consequences. Ten job incumbents completed the task rating questionnaire. A task importance value (see Levine, 1983) was computed for each task statement using the following formula: Task Importance Value = Time Spent + (Difficulty \times Criticality). While there is little empirical evidence that this computation of Task Importance is superior to others, its arithmetic demands that tasks be both critical and important to be considered important. If they are performed frequently, they take on added importance.

KSAs were rated by nine workers on three different dimensions: (1) the necessity for newly hired employees to possess the KSA (yes or no); (2) the extent to which trouble is likely if the KSA is ignored in selection (1-7 point scale); and (3) the extent to which the KSA distinguishes between superior and average workers (1-7 point scale).

Analysis of the Questionnaire

Interrater reliabilities of the task importance values and of the KSA ratings, as evidenced by coefficient alpha, were between .87 and .90 for all rating dimensions, with the exception of the "Distinguishes Superior and Average Workers" scale for which alpha was .72. Interrater reliability of ratings of individual pairs of job incumbent ratings ranged from approximately .40 to .70. Interrater reliability, as opposed to interrater agreement indices (Tinsley & Weiss, 1975), were calculated since the interest was in determining agreement as to the relative level of task and KSA importance.

For development of selection instruments, those KSAs that were rated by a clear majority as necessary for new workers and that received a high rating for trouble likely when ignored in selection were considered important selection criteria. These "important" KSAs were then categorized by the investigators and three expert raters into the following dimensions: communication skills; cooperativeness; emotional control; judgment; memory; and clerical/technical skills. Initial categorizations were done independently; discussion and revision of KSA statements resolved disagreements regarding the assignment of KSA statements to categories. In Table 1 we present the major KSA

TABLE 2

Examination Plan

Dimension	Oral directions	Interview	Simulation
Communication Skills		x	x
Emotional Control		x	x
Judgment		x	x
Cooperativeness		x	x
Memory	x		
Clerical/Technical Skills	x		

dimension and illustrative KSA statements.

Examination Plan

Both the KSAs and task statements played a role in the construction of selection tests. Each test devised centered around two or more of the KSA dimensions and focused on measuring the important aspects relevant to that dimension. An attempt was made to devise a test of a given KSA dimension that resembled the tasks for which the KSA was required. The tests constructed were based on tasks rated as most important and represented actual behaviors required on the job whenever possible. Tasks with Importance Values (Levine, 1983) over 18, which represented over one-half of the total number of tasks, became the source of examination materials. In the job analysis, experts generated KSAs relevant to each task. In test construction, the important tasks became item content by which to evaluate those KSAs. While no formal judgments of the relevance of each task were collected, content validity indices reported below support the assertion that the test items were representative of job content.

In order to adequately tap each of the KSA dimensions rated as critical to entry-level technicians, several selection tools were devised. The examination plan consisted of three consecutive phases: (1) an oral directions/typing test; (2) a situational interview; and (3) a telephone call simulation test. The matrix in Table 2 illustrates the dimensions tapped by each test.

The selection strategy involved multiple hurdles and both compensatory and noncompensatory scoring. Applicants who met or exceeded the specified cutoff scores in the oral directions/typing phase of the examination process proceeded with the situational interview and telephone call simulation, in which a compensatory scoring scheme was employed. The use of tests in this fashion was a result of both administrative and substantive considerations.

In addition to the usual concerns insuring that all proposed selection instruments be job-related, it was also necessary to meet the administrative requirements of the Center. While police personnel had previously interviewed candidates and wished continued input into the hiring decision, they wanted Civil Service personnel to screen most of the low qualified applicants, making the interview procedure less burdensome. Of course, the usual affirmative action demands were also presented and needed consideration. Because of these requirements, applicants proceeded through two phases of testing. Civil Service personnel administered an oral directions/typing test, which was designed for ease in administration to small groups of people. Police and supervisory personnel administered a situational interview and telephone call simulation.

Oral Directions/Typing Test

Description. The oral directions/typing test was designed to measure the applicant's memory ability and technical/clerical skills. This test consisted of four components: (1) spelling, (2) telephone call recordings, (3) monitoring, and (4) typing. The first three components of the exam were administered via a tape recorder; applicants listened to the information and questions presented on tape and responded in writing on answer forms.

The first part of the exam was a spelling test. A series of common street and place names not idiosyncratic to the local area were presented on tape, and applicants were required to write them down, spelling them correctly. Applicants were scored on the number of words spelled incorrectly.

The telephone call portion of the test matched tasks required for collecting and gathering information from callers. This portion comprised a series of 10 tape-recorded telephone call conversations between a caller and a police technician. In making this tape recording, the investigators listened to recorded telephone calls received by the emergency center. Calls involving a wide variety of incidents were collected, and the investigators wrote scripts of the telephone conversations, changing only the names and addresses of the people involved, when making these tape recordings for the test. The applicant listened to the calls and recorded pertinent information, such as the address of the incident, the caller's name, the nature of the incident, etc., onto a standardized form, which resembled the form technicians use when typing critical information from callers into the computer system. Applicants were given two practice calls to become acquainted with this procedure. Following each of the first five calls, applicants received

specific instructions telling them where to place each piece of information on the form. For the remaining five calls, applicants listened to the calls and recorded the information without specific instructions. Thus, applicants became familiar with the forms and how to use them before completing them without direction. Scoring was based on the accuracy of the information applicants recorded and on the amount of information applicants failed to record. These calls also served an important "training" function regarding the telephone call simulation test described below by acquainting applicants with the telephone procedure.

The third portion of the oral directions test was a monitoring exercise resembling tasks required for monitoring police units. Applicants listened to a series of statements that gave information about police units, their location, and activities. Following the presentation of these statements, questions were asked about the location of a unit, what type of call the officers were responding to, etc. Scores were based on the accuracy of the information recorded regarding the police units' locations and activities.

The final portion of the exam was a typing test. The typing test developed matched the kind of typing performed on the job. For this test, applicants typed information printed on standardized forms into blank forms of the type used in the telephone call portion of the test. Scoring was based on speed (the number of blanks typed in each form) and accuracy (the number of errors made while typing).

While the oral directions/typing test was designed to measure the KSA dimensions of memory and clerical/technical skills, the items in the test included actual job tasks rated as important. As an example, the monitoring exercise focused on the KSA dimension "memory ability" and was measured by focusing on the task of "monitoring the status of all field units knowing their activities and locations at all times."

Scoring scheme and cutoff scores. Each portion of the oral directions/typing test was scored separately. A cutoff score was set for each portion; applicants were required to meet or exceed the cutoff score for each portion separately as well as for the total test in order to continue with further testing.

In order to determine appropriate cutoffs for these tests, 25 undergraduate students at a large midwestern university were tested. While the investigators believed the students' test score distribution would be comparable to that of job applicants, arrangements were made to reevaluate the cutoff scores after test administration to a large pool of applicants. Administration of the test to a group of 75 candidates

yielded test score distributions that were nearly identical to those of the student pilot study.

For each of the four individual portions of the exam, especially the spelling and typing tests, the cutoffs were such that most people passed. In the case of the spelling exam, the job analysis indicated that some spelling ability was necessary in recording addresses and names, but that computer software to check spelling errors was available and being used. In the case of the typing exam, knowledge of the typewriter keyboard was essential, but accuracy was more important than speed in typing. Cutoffs for these four test portions—typing plus spelling plus telephone call recordings plus monitoring—had the effect of eliminating 43 of the original 75 candidates. These cutoffs were chosen for two reasons: (1) the feeling that minimal competence in these areas was necessary to a new job candidate's eventual success, and (2) the desire on the part of police personnel to evaluate only the best candidates with the more time-consuming selection interview and simulation.

The oral directions/typing test was scored by two independent raters to ensure that the scoring scheme and directions allowed objective and reliable scoring. Some minor problems concerning scoring directions were resolved, and subsequent interrater agreement in the scoring of the typing test exceeded 95% exact agreement for all rater pairs evaluated. Internal consistency reliability of the spelling, oral directions, and monitoring subtests were .78, .85, and .82 respectively, based on the scores of 75 applicants.

Situational Interview

Description. The situational interview, first proposed by Latham, Saari, Purcell, and Campion (1980), is a structured interview in which applicants respond to a series of job-related incidents. In developing our modification of situation interview questions for the police technicians' job, we began with the collection of critical incidents (Flanagan, 1954). Job incumbents were asked to identify and describe incidents in which particularly good or poor job behavior had been exhibited by some technician using Flanagan's criteria for incidents. After editing by the investigators, the incidents were translated into relevant interview questions in which job applicants were asked to indicate how they would behave in a particular situation. Since applicants were unlikely to have the job experience necessary to deal with the critical incidents, all critical incidents were translated into questions which job candidates would have some knowledge of and/or experience with.

For example, to demonstrate communication skill, the following critical incident was suggested by experienced workers:

A caller becomes abusive when talking to a technician. The technician gets mad and verbally abuses the caller using derogatory language.

This incident was transformed into the following interview question:

How would you react if you were a sales clerk, waitress, or service station attendant and one of your customers talked back to you, indicated you should have known something you did not, or told you that you were not waiting on them fast enough?

Seventeen interview questions were developed in order to evaluate the applicants on KSA dimensions of communication skills, emotional control, judgment, and cooperativeness. Since the interview questions were based on incidents of critical job behavior, tasks important to the job were the items in the interview questions.

Scoring scheme. A set of standardized BARS-type (Smith & Kendall, 1963) rating scales was devised to aid the interviewer-raters in making more objective judgments of applicants' performance in the situational interview. The interview rating guide consisted of a set of 12 scales used in evaluating candidates along job-relevant dimensions. On each scale, the dimension was listed along with a set of the interview questions that were likely to elicit candidate responses relevant to that particular dimension. Also listed were potential good and bad answers to each of the interview questions within each of the dimensions. Interview raters were instructed to review the interview questions relevant for that particular scale, consider the examples of good and poor answers to those questions, then make a rating on a 5-point scale ranging from excellent to poor, based on the applicant's responses. Ratings were made after all interview questions were asked and on the basis of notes taken during the interview because decision-making research on the interview (Arvey & Campion, 1982; Schmitt, 1976) suggests that this approach avoids recency and primacy effects. A total interview score for the applicant was computed by adding together the scores received on each scale.

In developing these rating scales, 25 undergraduate students were asked to respond to the interview questions twice, first as though they were job applicants and then as they imagined a "bad" job applicant would respond. Following this, two expert judges analyzed the responses and generated examples of good and poor answers to the questions, based on the responses provided by the students tested. These responses were transformed into the interview rating scales; one example of such is reproduced in Figure 1. The care with which interview

JUDGMENT/Common Sense: Ability to think and act clearly during emergencies.

In making your rating on this scale, consider answers to questions 9, 10, 11, 13, and 15.

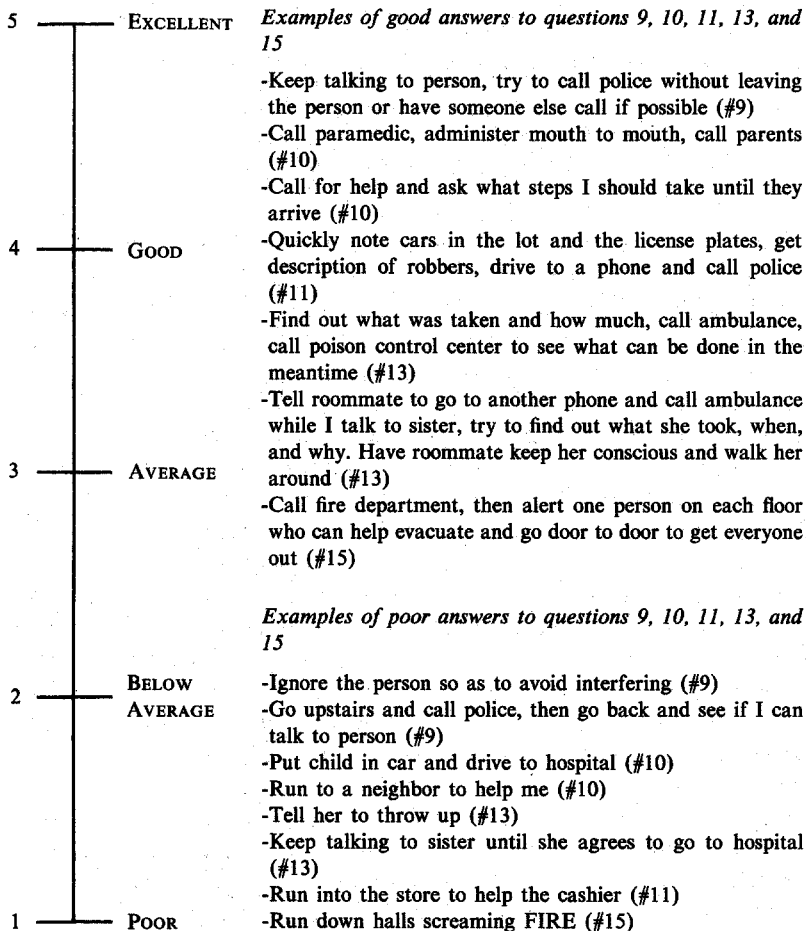


Figure 1. Example of rating scale developed for situational interview.

questions and rating scales were constructed resulted in high inter-rater reliability. Intercorrelations of the composite ratings provided by three interviewers on 25 candidates were .81, .82, and .89. The sum of the three raters' ratings had a coefficient alpha equal to .94.

CALLER: (Angry) I'm really mad and I've had enough of this. I have to work the morning shift and I get up really early. The morning shift starts at 6 o'clock. I work at Olds and I need my sleep. This is just ridiculous and it goes on every night. I can't believe it . . .

CBO Should interrupt at some point and ask what the problem is, get address, caller's name, etc.

CALLER: (Gives this information to caller *when asked*)

- name: Henry Abbott (or Harriete Abbott)
- incident: neighbor's dog barking furiously
- caller's address: 2728 W. Landsdown
- phone #: 378-5493
- city: Okemos
- neighbor's address: 2792 W. Landsdown

(when caller asks for this address, also say: I wish you'd hurry up and take care of this. I really need my sleep. You know, I have to work very early in the morning. This goes on every night and I'm tired of it. . .

CBO should interrupt and continue getting information by asking questions

- complaint against: don't know neighbor's name

Figure 2. Example of script for phone call simulation.

Telephone Call Simulation

Description. The telephone call simulation placed the job candidate in a role-playing exercise. The applicants played the role of a technician taking calls from complainants. An experienced technician played the role of a caller to the Center. This test was designed to assess the applicants' communication skills, emotional control, and judgment by focusing on the important tasks related to obtaining and recording critical information accurately.

Applicants talked with the caller, who was occasionally emotional or hysterical, and obtained the information necessary to send help to the caller. A variety of questions needed to be asked by the candidate in order to elicit the appropriate information. Candidates recorded information on a standardized form, the same form used during the oral directions/typing test.

The technician playing the caller was in the next room using special telephone equipment to speak with the applicant. This role-playing technician received three separate two-hour training sessions in which the importance of standardizing interaction with the candidates was stressed and interview scripts were refined. A series of six telephone call scripts (for an example, see Figure 2) were written to guide the technician in making calls. One practice call was given to allow applicants an opportunity to acquaint themselves with the procedure. The

telephone call conversations were tape-recorded so that raters could later listen to the tape and judge applicants' abilities. Tape-recording the conversations was useful because any non-standardized behavior on the part of the technician role-player could be taken into account and because it allowed multiple independent judgments of the candidates' performance. An anonymous reviewer of this paper suggested that the reliability of the responses of this role player should have been checked. This is certainly desirable, and such efforts were made during the role player's training; however, no determination of the constancy of the role players was made in the testing situation itself.

Scoring scheme. Scoring for the telephone call simulation consisted of two parts: (1) scoring of the standardized forms applicants completed and (2) ratings of the applicants' communication skills, emotional control, and judgment. The standardized forms were scored based on the amount of pertinent information obtained from the caller and on the accuracy of the information recorded on the forms.

A second score for the telephone call simulation was derived, based on raters' evaluations of the candidates' communication skills, emotional control, and judgment. As in the interview, raters listened to tape recordings of the telephone call simulation and made ratings on specially constructed BARS-type scales. The interrater correlations among the total scores, based on these scales, for the three raters were .69, .71, and .84; coefficient alpha for the sum of their ratings was .90.

Final evaluation of applicants. A total situational interview and telephone call simulation score was computed by totaling the scores on each of these exam portions. Job placement of the highest scoring individuals was recommended. While the Uniform Guidelines (1978) require further justification when a rank order strategy is used, our job analysis experts clearly indicated that the skills assessed in the telephone call simulation and interview were those that most clearly differentiated acceptable from superior performance. Unlike the typing and spelling skills, more skill on those dimensions was judged to be indicative of higher levels of performance.

In an effort to evaluate the degree to which the final examination components achieved our "behavioral consistency" goal and were content valid, a group of eight experienced technicians, who had helped in the construction of the exams, were asked to make judgments about the essentiality of each examination component. Specifically, they were asked to indicate the degree to which each exam provided information that was (1) essential, (2) necessary but not as essential as other questions, (3) useful, but not essential, or (4) not necessary or useful in judging applicants' skill in performing various aspects of the

job. Job dimensions were defined based on the tasks written for each dimension, and each exam component was judged against each task dimension. These judgments were summarized using Lawshe's (1975) content validity ratio:

$$CVR = [n_e - (N/2)] / (N/2)$$

where N equals the number of judges and n_e is the number of judges indicating an exam component was essential or necessary. For the interview, each interview question was judged separately and an average computed for the 17 interview questions. A summary of these judgments is contained in Table 3. Note that a 1.00 in this table indicates unanimous agreement on the part of the judges that an exam provided necessary or essential information about the candidates' ability to perform the job; -1.00 indicates unanimous agreement that an exam was not essential or necessary. This table indicates adequate content validity as judged by this set of experts with the exception of one task dimension: Dispatcher—Coordinate requests across jurisdictions. None of the examination elements contained items which represented samples of the tasks involved in this dimension. However, all components of the examination were perceived as relevant to one or more task dimension.

Discussion

The method of selection test development illustrated above exemplifies the behavioral consistency notion (Wernimont & Campbell, 1968), relying on consistencies between relevant dimensions of job behavior and pre-employment samples obtained in simulated job exercises. The emphasis was on important job-related behavioral measures defined through a thorough and systematic job analysis procedure. First, expert job incumbents were asked to generate a comprehensive list of task statements. They were then required to indicate what KSAs were required to perform each task. These KSAs were categorized, and tests were developed to measure these KSA dimensions. Those tasks for which a given KSA item was considered important became test items.

The simulations developed for these tests not only reproduced situations encountered on the job; they were designed so that inexperienced workers could respond. This type of procedure is advantageous in that it ensures job relevance and fairness in sampling job content. Tests devised concentrated only on those dimensions which new workers needed to possess. Further, skills and abilities which were necessary but required some training were modified to accommodate inexperienced workers. For example, the standardized forms used to record

information in the oral directions test and the telephone call simulation were modified and were used in written form rather than on a computer system; the situational interview questions were translated from critical job incidents to experiences a person could encounter in everyday life.

The behavioral consistency notion was reflected in several aspects of the selection procedure. The exam components matched, in so far as was possible, actual tasks required by the emergency telephone operator. For example, candidates were asked to (1) record information from actual calls on recording forms they would eventually use on the job; (2) spell words of streets and places in the city in which they worked; and (3) type onto the same forms which they would eventually use. In addition, the telephone calls in the oral directions and telephone call simulations were actual calls received by previous operators and were chosen to be representative of a wide range of situations. Interview questions were constructed in such a fashion as to represent situations similar to those encountered by the emergency operators though the behavior consistency of these questions may be less direct than other exam parts.

The development of these tests clearly relied on a content-oriented validation technique. The procedure used was in accordance with the standards established in the "Uniform Guidelines on Employee Selection Procedures" (1978). As stated in the Guidelines, a selection procedure can be validated by a content-oriented strategy if it is representative of the important aspects of performance on the job. The tests developed matched the important tasks performed on the job based on information provided in the job analysis. Further, the Guidelines specify that a job analysis which focuses on work behaviors and associated tasks is required. The job analysis procedure outlined above provided specific job behaviors as well as an assessment of their importance and criticality to the job, and test items were written to be similar to important job tasks.

The degree to which this content validity objective was achieved is summarized in Table 3. For the most part, the major task dimensions were evaluated by our test. However, as noted above, one area did not seem to be represented in the examination materials. While it is certainly likely that some of the tests, monitoring, for example, represent tasks that are similar to those involved in coordination of units across districts, no test item sampled directly the tasks in this component. While such items would have been a useful addition to the examination materials, very few tasks were part of this dimension, and tasks in the dimension as a whole occurred very infrequently.

TABLE 3

Summary of Content Validity Judgments

Major Task Dimension ^a									
	CBO	CBO	CBO	Dispatcher	Dispatcher	Dispatcher	Dispatcher	Lein	
	Collect info & ask ques- tions; speak clearly & calmly	Analyze info; type requests; transmit requests	Miscell- aneous duties (i.e., typing reports, record keeping)	Transmit requests, dispatch units, coordinate units, transmit info	Monitor units & emergencies, know loca- tion of units; provide info & back-ups; ask for info	Enter info about call and action	Coordinate requests across jurisdic- tions	respond to requests, transmit info, type, records, file info	
Oral directions	.75	1.0	.50	-.50	-.50	.25	-1.0	-.25	
Monitoring	0.00	0.00	-.25	0.00	1.0	-.50	-1.0	-.25	
Spelling	.25	.25	.50	-.25	-.75	.25	-1.0	.25	
Typing	.25	1.0	.75	-.75	-1.0	0.00	-.75	.75	
Phone call (written)	1.0	.50	0.00	-.25	-.25	0.00	-.75	.25	
Phone call (oral)	1.0	.50	-.50	.75	.25	-1.0	-.75	-.50	
Interview (average of 17 items)	.46	.53	-.82	-.10	-.12	-.78	-.75	-.60	

^aCBO (Complaint board operator), dispatcher and Lein operator refer to the three different positions occupied by the police technician. Workers rotated among these positions. In the content validity questionnaire, all major task dimensions were described in much more complete fashion.

The focus on standardization of test materials, the presentation of telephone calls, and the behavioral descriptions used for rating purposes were reflected in unusually high interrater reliabilities. Estimates of interrater reliability were usually in the area of .60; for both telephone call simulations and interviews, the reliability of a single rater ranged from .69 to .89, and composite ratings for three raters were in excess of .90.

A final advantage of this type of selection procedure is that it provides an extremely realistic preview of the job. Applicants gain a clear understanding of the job in question as they perform behaviors that are required on the actual job. Informal comments made to the test administrators underscored the impact this experience had on their perception of the emergency telephone operator's job. This "face validity" is important for two reasons cited by Dreher and Sackett (1983). First, it is argued that applicants are more motivated and react more favorably to face valid measures. Second, face valid measures are perceived as fairer and more appropriate (Sackett and Dreher, 1982) than other tests and are more likely to be used by decision makers.

From the authors' perspective, however, there are at least two problems with the simulations described. First, while the testing situation was relatively stressful for job candidates, actual emergency calls involving life-threatening situations will never be handled by an applicant in a testing situation. Second, there is ample evidence that minimal cutoff procedures do not optimize the validity or utility of a test (e.g., see Schmidt, Mack, & Hunter, 1984). Procedures for establishment of cutoffs are not particularly satisfying, perhaps because they represent an unrealistic assumption about the distribution of human abilities. We tried, based on the score distributions of a pilot group, and to some degree, succeeded in establishing cutoffs that would eliminate only those who performed very poorly on one or more of the test components. In retrospect, other judgmental procedures, such as those described by Livingston and Zieky (1982), may have provided more substantively meaningful and justifiable cutoff scores.

REFERENCES

- Arvey RD, Campion JE. (1982). The employment interview: A summary and review of recent research. *PERSONNEL PSYCHOLOGY*, 35, 281-322.
- Asher JJ, Sciarrino JA. (1974). Realistic work sample tests: A review. *PERSONNEL PSYCHOLOGY*, 27, 519-533.
- Dreher GF, Sackett PR. (1983). *Perspectives on employee staffing and selection*. Homewood, IL: Richard D. Irwin.

- Fine SA, Wiley WW. (1971). *An introduction to functional job analysis methods for manpower analysis* (Monograph No. 4). Kalamazoo, MI: W. E. Upjohn Institute.
- Flanagan JC. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Guion RM. (1974). Open a new window: Validities and values in psychological measurement. *American Psychologist*, 29, 287-296.
- Latham GP, Saari LM, Purcell ED, Campion MA. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Lawshe CH. (1975). A quantitative approach to content validity. *PERSONNEL PSYCHOLOGY*, 28, 565-575.
- Levine EL. (1983). *Everything you always wanted to know about job analysis*. Tampa, FL: Mariner Publishing Co.
- Livingston SA, Zieky MJ. (1982). *Passing scores*. Princeton, NJ: Educational Testing Services.
- Mento AJ. (1980). *The behavioral consistency selection procedure for entry level selection*. Paper presented at the American Psychological Association Convention, Montreal, Canada.
- Robertson IT, Kandola RS. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, 55, 171-183.
- Robinson DD. (1981). Content-oriented personnel selection in a small business setting. *PERSONNEL PSYCHOLOGY*, 34, 77-87.
- Sackett PR, Dreher GF. (1982). *Face validity and empirical validity as determinants of selection decisions*. Paper presented at the American Psychological Association Convention, Washington, DC.
- Schmidt FL, Caplan JR, Bemis SE, Decuin R, Dunn L, Antone L. (1979). *The behavioral consistency method of unassembled examining*. Washington DC: U.S. Office of Personnel Management.
- Schmidt FL, Mack MJ, Hunter JE. (1984). Selection utility in the occupation of U.S. park ranger for three modes of test use. *Journal of Applied Psychology*, 69, 490-497.
- Schmitt N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *PERSONNEL PSYCHOLOGY*, 29, 79-101.
- Smith PC, Kendall LM. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149-155.
- Tinsley HA, Weiss DJ. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22, 358-376.F
- Uniform guidelines on employee selection procedures. (1978). *Federal Register*, 43, 38290-38315.
- Wernimont PF, Campbell JP. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372-376.

Copyright of Personnel Psychology is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.