



Staffing Work Teams: Development and Validation of a Selection Test for Teamwork Settings

Michael J. Stevens
Purdue University

Michael A. Campion
Purdue University

The purpose of this research was to develop a selection test for staffing work teams. The knowledge, skills, and abilities (KSAs) for effective teamwork (Stevens & Campion, 1994) were used to develop a paper-and-pencil test of teamwork situations. KSAs reflected conflict resolution, collaborative problem solving, communication, goal setting and performance management, and planning and task coordination. Standard test construction techniques were used, and the test contained 35 multiple-choice items on hypothetical teamwork situations. Two validation studies were conducted involving production employees in a pulp mill ($n = 70$) and box plant ($n = 72$). Results showed criterion-related validity with supervisory and peer ratings of teamwork and overall job performance. However, a key unexpected finding was the large correlation with employment aptitude tests, suggesting that the Teamwork Test has a significant general mental ability component. The study also has secondary implications for the literatures on both contextual performance and situational tests. It was concluded that a consideration of individual level KSAs can have both conceptual and practical value in the staffing of work teams.

Although the study of work groups and group processes has a long and continuous history in organizational studies dating back more than half a century (e.g., Mayo, 1933; Trist & Bamforth, 1951; Walton, 1972), team-based work systems are experiencing a resurgence of interest in organizations today, and they are attracting substantial research attention as well (for recent example reviews see Goodman, 1986; Guzzo & Shea, 1992; Swezey & Salas, 1992). Yet, there is still much we do not know about many essential issues related to the management of work teams. This study addresses one such issue: How should we select

Direct all correspondence to: Michael J. Stevens, University of Missouri—St. Louis, School of Business Administration, 8001 Natural Bridge Road, St. Louis, MO 63121; Phone: (314)516-5000.

Copyright © 1999 by JAI Press Inc. 0149-2063

employees for teams? The study briefly summarizes the content domain of required knowledge, skills, and abilities (KSAs) needed to be an effective team member. Then, a paper-and-pencil selection test of teamwork situations is developed and validated in two separate studies.

Identification of the Teamwork KSA Content Domain¹

The conceptual framework of teamwork KSAs introduced by Stevens and Campion (1994) was used to define the content domain for the development of the paper-and-pencil test. This framework was focused in three ways. First, it stressed those attributes which are KSAs (i.e., learnable behaviors, mental abilities, etc.), rather than personality traits or dispositions. Although recent meta-analyses have demonstrated that personality-based selection can provide some degree of validity (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991), the history of success has been much greater with ability-based selection strategies (Hunter, 1986; Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). Second, the framework included only those attributes which are at the individual team member level of analysis, as opposed to the group or organizational levels. That is, the focus was on how to hire individuals for teams and not, for example, on how best to configure some optimal combination of members for a team. To be sure, the higher levels of analysis are certainly essential to managing teams effectively and may include such issues as social facilitation, team diversity, group norms and cohesion, company culture and reward systems, the nature and quality of supervision, and so on. However, these things have been the focus of previous research while the individual level of team management has remained relatively neglected. And third, the framework focused on teamwork KSAs, rather than taskwork (or technical-related) KSAs (see next paragraph). While task-related KSAs are certainly important for teamwork, they are not unique to team settings.

While this study's focus on differentiating between teamwork versus taskwork KSAs is not well developed by current theories of job performance, there are, however, several emerging and tangentially related areas of research which address this perspective. For example, Borman and Motowidlo (1993) draw an analogous distinction between contextual performance and task performance. They refer to contextual performance as those behaviors which support the social, psychological, and organizational context in which the work is performed. Thus, it indirectly includes teamwork behaviors. Also, recent work by Van Scotter and Motowidlo has identified a key facet of contextual performance which they call "interpersonal facilitation" (1996: 525). This facet includes interpersonal skills, maintaining good working relationships, helping others, and behaviors which appear to be direct aspects of teamwork performance. As another example, Organ (1988) distinguishes organizational citizenship behaviors (OCBs) from other aspects of required task performance. OCBs refer to extra-role and voluntary behaviors that help others perform their jobs and contribute to organizational effectiveness. Finally, Brief and Motowidlo (1986) have also defined a similar construct which they call prosocial organizational behavior. Teamwork activities

are commonly used as examples of these prosocial behaviors, even though organizations today may no longer consider these behaviors as extra-role or voluntary. Consequently, it might be viewed that this study's distinction between teamwork versus taskwork KSAs compares very directly to the kinds of distinctions being made between contextual performance and OCBs on the one hand, and task performance and in-role behavior on the other.²

The content domain of teamwork KSAs was identified through an extensive review of several major bodies of literature on groups, including organizational psychology, social psychology, socio-technical theory, and industrial engineering (see Stevens & Campion, 1994, for more details). These literatures do not explicitly identify individual level team member KSAs, however. Thus, this review did not simply summarize findings, but instead often inferred the individual level KSAs from the group and organizational level constructs. The review generated a content domain consisting of two main dimensions (i.e., interpersonal KSAs and self-management KSAs) with five sub-categories and 14 specific KSAs (see Table 1). While every effort was made to ensure that the review was thorough in terms of the range of KSAs identified, it must be recognized that the resulting framework may not necessarily be definitive.

Interpersonal KSAs

The interpersonal and social demands placed upon employees in team settings appear much greater relative to traditional individual-based work strategies. Thus, the impact of poor interpersonal skills are more consequential for teams. Work teams, therefore, appear to require employees who possess the following three sub-categories of interpersonal KSAs: (1) Conflict Resolution KSAs (i.e., the ability to recognize and encourage desirable, but discourage undesirable team conflict; to recognize the type and source of conflict and implement an appropriate resolution strategy; and to use integrative, rather than distributive, approaches to negotiation); (2) Collaborative Problem Solving KSAs (i.e., the ability to match the proper degree of participation to the problem; and to recognize obstacles to collaborative problem solving and implement appropriate corrective actions); and (3) Communication KSAs (i.e., the ability to recognize and utilize decentralized networks to enhance communication; to communicate openly and supportively; to listen nonevaluatively and use active listening techniques; to match one's own nonverbal and verbal messages and to recognize and interpret the nonverbal messages of others; and to engage in small talk and ritual greetings).

Self-management KSAs

Frequently in team-based interventions, control over many essential managerial or supervisory functions associated with the team's operations are turned over to the group. That is, the team is empowered to be self-managing. As such, the team members should possess the KSAs required to perform these basic managerial or supervisory duties. The two sub-categories of self-management KSAs used in this study are: (4) Goal Setting and Performance Management KSAs (i.e., the ability to help establish specific, challenging, and accepted team goals; and to

Table 1. The Knowledge, Skill, and Ability (KSA) Requirements for Teamwork

-
- I. Interpersonal KSAs
 - A. Conflict Resolution KSAs
 - 1. The KSA to recognize and encourage desirable, but discourage undesirable team conflict.
 - 2. The KSA to recognize the type and source of conflict confronting the team and implement an appropriate resolution strategy.
 - 3. The KSA to employ an integrative (win-win) negotiation strategy, rather than the traditional distributive (win-lose) strategy.
 - B. Collaborative Problem Solving KSAs
 - 4. The KSA to identify situations requiring participative group problem solving and to utilize the proper degree and type of participation.
 - 5. The KSA to recognize the obstacles to collaborative group problem solving and implement appropriate corrective actions.
 - C. Communication KSAs
 - 6. The KSA to understand communication networks, and to utilize decentralized networks to enhance communication where possible.
 - 7. The KSA to communicate openly and supportively, that is, to send messages which are (a) behavior- or event-oriented, (b) congruent, (c) validating, (d) conjunctive, and (e) owned.
 - 8. The KSA to listen nonevaluatively and to appropriately use active listening techniques.
 - 9. The KSA to maximize the consonance between nonverbal and verbal messages and to recognize and interpret the nonverbal messages of others.
 - 10. The KSA to engage in small talk and ritual greetings and a recognition of their importance.
 - II. Self-management KSAs
 - D. Goal Setting and Performance Management KSAs
 - 11. The KSA to help establish specific, challenging, and accepted team goals.
 - 12. The KSA to monitor, evaluate, and provide feedback on both overall team performance and individual team member performance.
 - E. Planning and Task Coordination KSAs
 - 13. The KSA to coordinate and synchronize activities, information, and tasks between team members.
 - 14. The KSA to help establish task and role assignments for individual team members and ensure proper balancing of workload.
-

Notes: From "The Knowledge, Skill, and Ability Requirements for Teamwork: Implications for Human Resource Management" by M. J. Stevens and M. A. Campion, 1994, *Journal of Management*, 20, p. 505. Copyright 1994 by JAI Press. Reprinted by permission

monitor, evaluate, and provide feedback on performance); and (5) Planning and Task Coordination KSAs (i.e., the ability to coordinate and synchronize activities, information, and tasks between team members; and to help establish task and role assignments for individual team members and ensure proper balancing of workloads).

Test Development and Hypotheses

Test Development

The content domain described above was used to guide the development of a paper-and-pencil selection test (hereafter called the "Teamwork Test"). The test focused on teamwork knowledge because of the long history of success in using

knowledge-based tests for selection, and because of the role of job knowledge in path models of performance (Borman, 1991; Dye, Reck, & McDaniel, 1993; Hunter, 1986). Standard test construction procedures were followed to develop the appropriate test items (Anastasi, 1988; Nunnally, 1978).

Situational questions were developed for the test, wherein hypothetical teamwork situations were depicted. Situational questions were used for three reasons. First, such questions have shown high validity in other selection instruments, especially structured interviews (M. Campion, J. Campion, & Hudson, 1994; File & Remmers, 1946; Latham, Saari, Pursell, & Campion, 1980; Motowidlo, Dunnette, & Carter, 1990; Weekley & Gier, 1987). Second, when written in terms of job situations, such questions can have high face validity. And third, there is some evidence from the research on tacit intelligence that such questions may allow for the measurement of attributes independent of general mental ability (e.g., Sternberg, Wagner, Williams, & Horvath, 1995), although recent meta-analytic evidence suggests that situational tests are commonly related to mental ability (McDaniel, Finnegan, Morgeson, Campion, & Braverman, 1997). Therefore, another contribution of this study is to add to the literature on situational tests, both in terms of additional validity evidence and in terms of assessing whether they measure constructs different than ability.

Respondents indicated what they would do in each situation by selecting from four multiple-choice alternatives. The alternatives were written to have similar social desirability to reduce faking. The correct alternative was determined from findings in the research literature. Questions were dichotomously scored (0/1) and total scores were the sums of correct answers.

Initially, 46 questions were written and pilot tested on 234 undergraduate students at a major midwestern university. This number met the minimum suggested subject-to-item ratio of five-to-one required for reliable test item analysis (Nunnally, 1978). Items were eliminated or revised based on difficulty, discriminability, and item-total correlations. In addition, a reading level assessment was conducted (Flesch, 1948; Fry, 1968). The final version of the Teamwork Test consisted of 35 items with an eighth grade reading level requirement (see Appendix for sample items).

Content validity of the Teamwork Test was verified by having four independent judges assign questions to the five sub-categories in Table 1. There was 100% agreement on 28 questions and 75% agreement on 6 questions. Results showed 4 questions reflected conflict resolution, 8 reflected collaborative problem solving, 11 reflected communication, 5 reflected goal setting and performance management, and 6 reflected planning and task coordination.

Hypotheses

Hypotheses were developed to assess the convergent, criterion-related, and incremental validity of the Teamwork Test.

Convergent Validity

If the Teamwork Test is measuring the intended constructs, then it should converge in a predictable fashion to other known measures of the same or related

constructs (Schwab, 1980). The Teamwork Test may have a large mental ability component because it is a knowledge test and uses a paper-and-pencil format. As such, a high degree of convergence would be expected with traditional employment aptitude tests based on the positive correlations observed among mental ability tests (e.g., math, verbal, mechanical ability; Jensen, 1986). Thus,

H1: The Teamwork Test should correlate positively with traditional employment aptitude tests.

Criterion-Related Validity

Consistent with the individual level focus of the present study, the Teamwork Test should show criterion-related validity with job performance of individual employees. Thus,

H2: The Teamwork Test should correlate positively with measures of individual team member performance.

Further, because the emphasis is on teamwork rather than taskwork performance, the Teamwork Test should be more highly correlated with teamwork performance than should the employment aptitude tests, whereas the aptitude tests should be more highly correlated with taskwork performance than should the Teamwork Test. Thus, the following hypotheses not only address criterion-related validity, but also construct validity in that they help explain the underlying relationships between the predictors and the various constructs of performance.

H3a: The correlation between the teamwork performance ratings and the Teamwork Test should be greater than the correlation between the teamwork performance ratings and the employment aptitude tests.

H3b: The correlation between the taskwork performance ratings and the employment aptitude tests should be greater than the correlation between the taskwork performance ratings and the Teamwork Test.

Incremental Validity

To demonstrate that the Teamwork Test contributes additional or incremental value to the body of existing selection measures, it must be able to add predictive validity to selection systems beyond that which is obtainable from readily available alternative measures. Because traditional employment aptitude tests are widely used, readily available, and among the best overall predictors of performance, they would, thus, constitute an appropriate standard of comparison for a new selection instrument. Thus,

H4: The Teamwork Test should have incremental criterion-related validity beyond the employment aptitude tests.

Study One

Sample

Study 1 was conducted at a southeastern pulp mill. The company was replacing its pulp mill and needed to staff the new facility. The new mill was to have a management approach based upon work teams and self-management. All subjects ($n = 70$) were current employees applying for the new jobs. Statistical power was 82% to detect an $r = .30$ ($p < .05$, one-tailed test due to directional predictions; results at the marginal probability level [i.e., $p < .10$] will also be interpreted to enhance statistical power given the sample and effect sizes; Cohen, 1977).

Method

The setting provided three advantages for the present study. First, because test performance would be used to determine who received the new jobs, test taking motivation was very high (Arvey, Strickland, Drauden, & Martin, 1990). Second, current employees had originally been hired without any ability-based testing or valid interviews, all promotions had been driven by seniority only, and there had been no systematic differences in training given to employees. Thus, restriction of range on tested abilities did not seem likely. Third, a concurrent validation design was possible because measures of current job performance could be obtained, and a job analysis revealed that the team-related attributes required by the new jobs were also required on the current jobs (e.g., interpersonal communication, group problem solving, employee coordinating and synchronizing of interdependent work activities, self-direction and evaluation of performance, etc.).

Predictors

Teamwork Test. The Teamwork Test was administered to all employees. Two of the 35 item-total correlations were negative, and those items were therefore eliminated from subsequent analyses. For the remaining 33 items, the alpha internal consistency reliability was .80.

Employment Aptitude Tests.³ Nine traditional employment aptitude tests were included to provide an extensive evaluation of the abilities typically assessed in employment settings. Three measured verbal ability: Ramsay Job Skills Reading Test (Form A) measured reading comprehension (Ramsay, 1991); Personnel Test for Industry (Verbal, Form A) measured vocabulary (Wesman, 1980); and Flanagan Industrial Test (Expression) measured expression/grammar (Flanagan, 1975). Three measured quantitative ability: Science Research Associates' Arithmetic Index measured basic arithmetic (Science Research Associates, 1986); Psychological Services Incorporated Basic Skills Tests (Problem Solving, Form A) measured math problem solving (Ruch, Weiner, McKillip, & Dye, 1985); and Flanagan Industrial Tests (Scales) measured the ability to read scales and graphs (Flanagan, 1975). Two measured perceptual ability: Employee Aptitude Survey (#3, Form A) measured visual pursuit (Ruch & Ruch, 1980); and Employee Aptitude Survey (#4, Form A) measured visual speed/accuracy (Ruch & Ruch, 1980). The final test measured mechanical ability: Differential Aptitude Test measured mechanical reasoning (Bennett, Seashore, & Wesman, 1991).

Tests were administered according to publisher instructions and scored in terms of number correct. Because internal consistency reliability measures are inappropriate for speeded tests, they were not computed. However, where other forms of reliability were reported in test manuals, the levels were acceptable (e.g., alternate forms, test-retest, typically ranging from the .70s to the .90s). An overall *aptitude test composite* score was calculated by converting the nine individual test scores to *z*-scores and then averaging. Alpha internal consistency reliability applied to this 9-test composite was .93 (which may be an underestimate of the reliability of a multi-test composite; Cortina, 1993; Wherry & Gaylord, 1943).

Criteria

Criterion measures were supervisory ratings obtained as input for the new plant staffing decisions and not part of an on-going company appraisal process. Supervisors were instructed that ratings would be kept strictly confidential and not fed back to employees or included in personnel files.

Five items were created for the study: three reflected *teamwork performance* (self-management, team contribution, and communication), and two reflected *taskwork performance* (technical knowledge and learning orientation). An *overall performance* measure was also obtained by combining all items. To assess the appropriateness of separating the performance criteria into both the teamwork and taskwork dimensions, a confirmatory factor analysis was conducted (with data combined from both studies to attain an adequate sample size). Results demonstrated that a two-factor job performance model fit the data better than a one-factor model (change in chi-square = 18.89, $df = 1$, $p < .05$; goodness-of-fit index = .94 versus .89).⁴

Five supervisors provided ratings on all employees in their areas who applied for the new jobs, with each supervisor rating 30 employees on average. To enhance reliability, ratings were provided by multiple independent supervisors. Usually two (59% of cases) or three (27%) supervisors provided ratings for each employee. Supervisors were trained on the rating instrument and process. Items were defined by a brief explanation and coupled with a 5-point scale ranging from 5—"well above average" (top 20% of employees) to 1—"well below average" (bottom 20% of employees).

Criterion scores were simply computed as averages across the relevant items from the rating forms. However, in order to eliminate any potential response biases from each of the supervisor's evaluations (e.g., varying amounts of leniency or severity), their ratings were first standardized by converting to *z*-scores. Then, subjects' scores for individual items on the scale were generated by averaging the standardized ratings across the supervisors. Finally, composite scores for taskwork and teamwork performance were created by combining (i.e., averaging across) the relevant items on the form. Alpha internal consistency reliability was .87 for the three teamwork performance items, .86 for the two taskwork performance items, and .90 for all five items combined. The interrater reliability (intraclass correlation) of the mean of the raters was .84 for the overall composite, and from .74 to .83 for the individual items. The average absolute interrater differences were .54 *SD* for the overall composite, and from .66 to .81

SD for individual items. Thus, supervisor ratings showed acceptable reliability and agreement.

Results

Descriptive statistics and intercorrelations are shown in Table 2. The data were compared to the norms in the manuals for the aptitude tests, and the means and distributions were examined for the other measures, and no obvious signs of ceiling or range restriction problems were found. Average education was 11.9 years ($SD = 1.6$), and average company tenure was 21.7 years ($SD = 8.3$). Company tenure showed moderate correlations with all three performance measures ($r = .29$ to $.33$; $p < .05$), but it showed no significant correlation with either the Teamwork Test or the battery of aptitude tests.⁴

Convergent Validity

Table 2 shows a positive correlation between the Teamwork Test and the employment aptitude test composite, thus supporting Hypothesis 1. The correlation of .81 was much larger than expected, however, given that the Teamwork Test was developed based on an analysis of the teamwork literature and not on traditional mental abilities such as reading and math.

Criterion-Related Validity

Table 2 provides confirmation of Hypothesis 2 by showing that the Teamwork Test correlates with ratings of teamwork performance ($r = .44$), taskwork performance ($r = .56$), and overall performance ($r = .52$). As predicted, the correlation between teamwork performance and the Teamwork Test is larger than the correlation between teamwork performance and the employment aptitude tests ($r = .44$ versus $.33$), but this difference is only significant at $p < .10$ ($t = 1.62$, $df = 67$, one-tailed), thus providing marginal support for Hypothesis 3a. Though in the expected direction, the correlation between taskwork performance and the aptitude test composite is not significantly larger than the correlation between taskwork performance and the Teamwork Test ($r = .60$ versus $.54$; $t = 0.67$, $df = 67$, ns), thus not supporting Hypothesis 3b.

Incremental Validity

To test for incremental validity, stepwise regression analyses were conducted for each of the three criteria (Table 3). There is a significant increase in explained variance by the Teamwork Test beyond the aptitude test composite for both teamwork performance (incremental $R^2 = .08$) and overall job performance (incremental $R^2 = .06$), but not for taskwork performance (incremental $R^2 = .01$). Thus, Hypothesis 4 is largely supported. In addition, the incremental validity for the Teamwork Test in predicting teamwork performance offers supplemental support for Hypothesis 3a, while the incremental validity of the aptitude test composite in predicting taskwork performance offers supplemental support for Hypothesis 3b.

These incremental validity results are noteworthy given the high correlation of .81 between the Teamwork Test and the aptitude tests (see Hypothesis 1). Depending upon which reliability estimates are used (e.g., internal consistency or

Table 2. Descriptive Statistics and Intercorrelations for Study 1 Measures

Measures	M	SD	Intercorrelations														
			1	2	3	4	5	6	7	8	9	10	11	12	13		
1. Teamwork Test	19.8	5.5															
2. Reading Comprehension	30.9	8.1	77														
3. Vocabulary	27.1	10.5	77	81													
4. Expression/Grammar	11.8	4.1	65	74	78												
5. Basic Arithmetic	34.3	12.0	70	74	71	64											
6. Math Problem Solving	7.5	3.7	70	70	73	69	78										
7. Scales and Graphs	8.2	5.0	65	61	65	56	60	63									
8. Visual Pursuit	14.0	6.0	55	54	57	45	54	55	59								
9. Visual Speed/Accuracy	79.0	27.5	52	50	62	55	63	64	50	53							
10. Mechanical Reasoning	30.6	7.1	65	69	64	54	55	57	58	56	34						
11. Aptitude Test Composite	0.0	0.8	81	87	89	82	85	86	78	73	75						
12. Teamwork Performance	0.0	0.7	44	38	28	24	28	27	28	32	14 ^a	33					
13. Taskwork Performance	0.0	0.8	56	56	57	55	43	49	46	50	34	46	72				
14. Overall Performance	0.0	0.7	52	49	43	40	37	38	38	42	24	35	47	95	90		

Notes: *n* = 70. Decimals omitted from correlations. All correlations are significant at least at *p* < .05, except the correlation indicated with superscript "a" which is not significant.

Table 3. Regression Analyses Assessing Incremental Validity of the Predictor Measures in Study 1

<i>Measures</i>	R^2	ΔR^2	b^a
Teamwork Performance (Total $R^2 = .19^{**}$):			
Teamwork Test	.19	.08	0.48**
Aptitude Test Composite	.11	.00	-0.06
Taskwork Performance (Total $R^2 = .37^{**}$):			
Teamwork Test	.31	.01	0.21
Aptitude Test Composite	.36	.06	0.42**
Overall Performance (Total $R^2 = .28^{**}$):			
Teamwork Test	.27	.06	0.40**
Aptitude Test Composite	.22	.01	0.15

Notes: $n = 70$. $df = 2/67$. R^2 = variance explained when variable is entered first. ΔR^2 = incremental change in R^2 when variable is entered last. b^a = standardized regression coefficient for final equation. ** $p < .01$.

interrater), correcting for both tests yields an estimated correlation ranging from .97 to .99. Thus, the tests would appear to be redundant, which would seem to preclude incremental changes in R^2 . However, with less than perfectly reliable tests, it may be possible that the incremental changes are due simply to adding more reliability to the measurement of the single shared construct, rather than to measuring a different construct (Aiken & West, 1991; Gordon, 1968).

Study Two

Sample

Study 2 took place at two northeastern cardboard box plants. The nature of the work at the two plants was essentially identical. Unlike Study 1, participation was voluntary and not linked to any work-related outcomes. One plant provided 34 subjects, of which all but 2 were nonexempt (hourly) employees. The other plant provided 38 subjects, of which 19 were nonexempt and 19 were exempt (salaried) employees. This provided a combined sample of $n = 72$ subjects for Study 2, for an overall participation rate of 46% at the two plants. Analyses showed that plant and exempt status did not explain any variance in the criteria beyond the predictors, so all subsequent analyses were conducted with both plants and all employees combined. Statistical power was 83% to detect an $r = .30$ ($p < .05$, one-tailed; results at the marginal probability level [i.e., $p < .10$] will also be interpreted to enhance statistical power given the sample and effect sizes; Cohen, 1977).

Method

Study 2 was conducted for three reasons. First, it provided a replication of Study 1 with a new sample. This was desirable given the small sample size in

Study 1. Second, it was believed that the setting for Study 2 was relatively farther along in the transition to a team-based work system (e.g., in terms of its usage of quality circles and problem-solving task groups), and thus provided a test of the hypotheses in a setting which was more advanced in terms of teamwork requirements (e.g., collaborative problem solving, communication demands, etc.). And third, Study 2 allowed for expanded criteria by using peer and self-ratings in addition to supervisory ratings.

Predictors

Teamwork Test. Subjects were administered the identical 35-item version of the Teamwork Test used in Study 1. One of the two items with a negative item-total correlation in Study 1 also had a negative correlation in Study 2 and was eliminated. The other item was also eliminated to maintain consistency across studies, but no differences in results occurred either with or without this second item. For the remaining 33 items, the alpha internal consistency reliability was .81.

Employment Aptitude Tests. Due to time constraints, only two of the nine employment aptitude tests from Study 1 were used: *vocabulary* and *math problem solving*. These tests were chosen because they had the highest criterion-related validities in Study 1, they represented the most common constructs in traditional selection test batteries (i.e., verbal and math), and they were efficient in terms of time usage. An overall *aptitude test composite* score was calculated by converting individual test scores to *z*-scores and then averaging. The alpha internal consistency reliability was .74, which is lower than that in Study 1 because fewer tests were included. Again, this may be a slight underestimate of the reliability of a multi-test composite (Wherry & Gaylord, 1943).

Criteria

Supervisory Ratings. Supervisory ratings were obtained specifically for the validation study and not part of an on-going company appraisal process. Supervisors were instructed that ratings were for research purposes only, and that all information would be kept strictly confidential and not fed back to the employees or kept in personnel files. The measure was a slightly expanded version of the ratings used in Study 1. The supervisor teamwork rating included five items (resolving conflicts, collaborative behaviors, interpersonal communication, goal setting and performance management, and coordinating and planning). The supervisor taskwork rating included three items (technical knowledge depth, technical knowledge breadth, and learning orientation). A supervisor overall rating was also obtained by combining all items. As described in Study 1, a confirmatory factor analysis of the supervisory ratings supported a two-factor model.

Twenty supervisors provided the ratings on all employees in their areas who participated in the study. To enhance reliability, ratings were provided by multiple independent supervisors. The number of supervisors was two (22% of cases), three (57%), four (15%), or five (6%) for each employee. Rating scales, standardization, and other procedures were the same as in Study 1.

Alpha internal consistency reliability was .95 for the five teamwork performance items, .85 for the three taskwork performance items, and .95 for the eight items combined. The interrater reliability (intraclass correlation) of the mean of the raters was .84 for the overall composite, and from .71 to .82 for individual items. The average absolute interrater differences were .59 *SD* for the overall composite, and from .60 to 1.06 *SD* for individual items. Thus, supervisory ratings showed acceptable reliability and agreement.

Self-Ratings. Subjects completed a self-rating form that was identical to the supervisor form, but with wording changed to the first-person. Employees were instructed to provide “descriptions” of their own work behaviors relative to their co-workers. They were told that this information was for research purposes only, and that all responses would be kept strictly confidential and not shown to management or placed in personnel files. Alpha internal consistency reliability was .83 for teamwork performance, .63 for taskwork performance, and .81 for overall performance.

Peer Nominations. Peer nominations were used rather than peer ratings because of their presumed higher validity and reliability (Kane & Lawler, 1978), and because the unionized setting discouraged use of a measure that might ask employees to make negative ratings of other employees. The nomination form contained names of co-workers from the same shift who were also participating in the study. It asked the nominator to identify the top one-third preferred peers in each of three categories: (1) “For promoting good working relationships;” (2) “For helping take charge and staying focused;” and (3) “For technical expertise and know-how.” A “Do not know this person well enough” option was also allowed. These categories were chosen to reflect the interpersonal, self-management, and task performance dimensions, respectively. Peers were instructed that they could nominate the same person for more than one category, that everyone did not have to be nominated, and that their responses were for research purposes only and would be held strictly confidential (e.g., co-workers, supervisors, or members of management would not see the responses).

Because subjects differed in terms of the number of possible times they could be nominated, scores were calculated as relative percentages (i.e., the total number of times a person was nominated divided by the total number of times the person could have been nominated). Categories (1) and (2) were averaged to obtain the *peer teamwork nomination*, while Category (3) provided the *peer taskwork nomination*. A *peer overall nomination* was obtained by averaging all three categories.

Results

Descriptive statistics and intercorrelations for Study 2 are shown in Table 4. The data were compared to the norms for the aptitude tests, and the means and distributions were examined for the other measures, and no obvious indications of ceiling or range restriction problems were found. Average education was 12.6 years (*SD* = 1.4), and average company tenure was 5.6 years (*SD* = 6.5). Company

tenure showed no significant correlation with any of the performance measures or the predictors.⁴

Convergent Validity

In support of Hypothesis 1, Table 4 shows a positive correlation between the Teamwork Test and the employment aptitude test composite ($r = .81$). This correlation is the same size as found in Study 1.

Criterion-Related Validity

Table 4 also shows that Hypothesis 2 is partially supported by positive correlations between the Teamwork Test and supervisory ratings of teamwork performance ($r = .21$), taskwork performance ($r = .25$), and overall performance ($r = .23$), as well as for peer nominations of teamwork ($r = .23$) and overall performance ($r = .21$). However, the Teamwork Test does not show criterion-related validity for peer nominations of taskwork performance nor for any of the self-ratings, although the self-ratings did follow the patterns suggested elsewhere in the literature of usually being less predictable than ratings provided by others (Harris & Schaubroeck, 1988; Shore, Shore, & Thornton, 1992).

Contrary to Hypothesis 3a, the correlations between teamwork performance and the Teamwork Test are not significantly greater than the correlations between teamwork performance and the employment aptitude tests. There is a slight reversal for the supervisory ratings ($r = .21$ versus $.23$), and the difference is not significant for the peer nomination ($r = .23$ versus $.19$; $t = 0.55$, $df = 62$, ns).

The findings in Table 4 suggest that Hypothesis 3b is partially supported. The correlation between taskwork performance and the aptitude test composite is significantly greater than the correlation between taskwork performance and the Teamwork Test for both the supervisory ratings ($r = .36$ versus $.25$, $t = 1.60$, $df = 69$, $p < .10$) and for the self-ratings ($r = .17$ versus $.07$, $t = 1.68$, $df = 63$, $p < .05$).

Incremental Validity

Stepwise regression analyses were used to assess incremental validity for Study 2. Contrary to Hypothesis 4, the Teamwork Test did not show significant incremental validity beyond the aptitude test composite for any of the criterion measures (for the teamwork criterion, incremental R^2 was $.01$ for the supervisory rating, ns; $.02$ for the self rating, ns; and $.01$ for the peer nomination, ns). Conversely, the aptitude test composite did show incremental validity beyond the Teamwork Test for the taskwork criterion measure for both the supervisory ratings (incremental R^2 was $.07$, $p < .01$) and self ratings (incremental R^2 was $.03$, $p < .10$).

Supplemental Analysis

Due to the relatively smaller sample sizes in both Studies 1 and 2, the data were combined ($n = 142$) to more accurately estimate the population validities. To maintain consistency, the combined analyses were conducted with only the two aptitude tests and five supervisory rating items common to both studies. Mean test scores did not differ between samples for either the Teamwork Test ($t = 0.92$, $df =$

Table 4. Descriptive Statistics and Intercorrelations for Study 2 Measures

Measures	M	SD	Intercorrelations														
			1	2	3	4	5	6	7	8	9	10	11	12			
1. Teamwork Test	18.3	5.8															
2. Vocabulary	30.8	10.6	81**														
3. Math Problem Solving	9.3	3.2	63**	58**													
4. Aptitude Test Composite	0.0	0.9	81**	89**	89**												
5. Supervisor Teamwork Rating	0.0	0.7	21*	18†	23*	23*											
6. Supervisor Taskwork Rating	0.0	0.7	25*	31**	33**	36**	79**										
7. Supervisor Overall Rating	0.0	0.7	23*	24*	28**	29**	97**	91**									
8. Self Teamwork Rating	4.0	0.7	-12	04	-11	-04	26*	25*	27**								
9. Self Taskwork Rating	3.8	0.7	07	18†	12	17†	-02	11	03	48**							
10. Self Overall Rating	3.9	0.6	-06	11	-03	04	18†	23*	21*	93**	77**						
11. Peer Teamwork Nomination	0.4	0.2	23*	14	19†	19†	25*	24*	26*	18†	02	14					
12. Peer Taskwork Nomination	0.3	0.3	-04	13	12	14	18†	27**	23*	21*	10	20†	17†				
13. Peer Overall Nomination	0.4	0.2	21*	18†	21*	22*	28**	33**	32**	25*	08	22*	83**	68**			

Notes: n = 72 for supervisor ratings, n = 66 for self ratings, and n = 65 for peer nominations. Decimals omitted from correlations. †p < .10, *p < .05, and **p < .01.

140, ns) or the aptitude test composite ($t = 1.02$, $df = 140$, ns). Results show that criterion-related validities for the Teamwork Test (aptitude test composite) of .32 (.27) for teamwork performance, .40 (.47) for taskwork performance, and .37 (.36) for overall performance. In addition, the intercorrelation between the Teamwork Test and the employment aptitude battery was somewhat lower ($r = .74$), and the Teamwork Test showed significant incremental validity beyond the aptitude test composite for predicting both teamwork performance (incremental $R^2 = .03$) and overall performance ratings (incremental $R^2 = .03$).⁴ However, as was mentioned previously, it is not known how much of this incremental prediction may be due to less than perfect measurement reliability.

Discussion

Summary and Conclusions

This study attempted to develop a paper-and-pencil test of teamwork situations for the purpose of personnel selection. It started with a fairly extensive review of the literature to determine the appropriate content domain of knowledge, skills, and abilities required for teamwork. Then, based on this content domain, standard test construction procedures were followed to develop a 35-item test, which was then validated in two separate studies. Criterion-related validity was found, but most surprising was the finding that the Teamwork Test correlated so strongly with a battery of traditional employment aptitude tests. When corrected for measurement error, estimates of the true correlation range from .91 to .99 (depending on whether data from the two studies are combined or examined separately), forcing the conclusion that it is a highly redundant measure of general mental abilities.

Even though it may be redundant, could the Teamwork Test still be useful for team selection? Yes, since it showed criterion-related validity in both studies for supervisory ratings of both teamwork and taskwork performance, and it also showed validity in Study 2 for peer nominations of teamwork. The Teamwork Test even demonstrated incremental validity beyond the aptitude tests, but it is unknown whether this is due to the less than perfect reliability of the measures. As such, these validities would tend to support the use of the Teamwork Test for guiding work team staffing decisions. In addition, the content of the Teamwork Test would likely be viewed as more face valid than traditional employment aptitude tests. Thus, for organizations with work teams, it may be an effective way to get information about applicants' mental abilities in a more face valid way than traditional tests.

Do the findings of this study mean that teamwork KSAs are nothing more than general mental ability? No, since this particular operationalization of teamwork KSAs undoubtedly predisposed a high relationship with mental ability tests. This was perhaps due in part to the fact that mental abilities, like reading ability and ability to think in verbal terms, were needed by the paper-and-pencil format of the Teamwork Test. This may be viewed as a form of common method variance. In addition, the situations posed by the test questions also required some amount of problem solving, which again reflects mental ability. Lastly, the Team-

work Test is a knowledge test that was developed to measure a specific content domain of team-related KSAs, rather than personality traits or dispositions; thus, it largely reflects general mental abilities (which also helps explain the strong correlation with ability tests). If future research should find that other operationalizations of teamwork KSAs are also related to general mental abilities, then one interesting implication is that the use of teams may require significant cognitive abilities on the part of team members and not just the capacity to be congenial or sociable with other team members. This would seem reasonable to expect in view of the fact that many team-based interventions are often implemented in conjunction with managerial change efforts emphasizing re-engineering, quality advances, continuous improvement, or new technologies. Such environments would tend to place a premium on the ability of employees to learn and innovate. Indeed, Hunter suggests that this may be the reason why general mental ability is such an effective predictor of performance, because "it measures the ability to innovate and prioritize in dealing with situations that deviate from those encountered" in prior experience (1986: 358).

This study also makes secondary contributions to two other literatures. First, it helps to further elaborate on the constructs identified in the contextual performance literature. As was noted earlier, teamwork may be viewed as an aspect of contextual performance, and this study helps define the KSAs required for teamwork performance, as well as providing a measure of those KSAs. In addition, this study also provides modest evidence for the empirical distinctiveness and differential predictability of the teamwork versus taskwork performance dimensions. Second, this study also adds further data to the mounting validity evidence supporting situational tests. Although some research (e.g., on tacit knowledge; Sternberg et al., 1995) has reached a different conclusion, this study clearly suggests that some situational tests, such as the Teamwork Test, may indeed be tapping constructs highly similar to mental ability.

Limitations and Future Research

The most significant limitation of the present study is that it developed a measure which had an unexpectedly high relationship with a battery of traditional employment aptitude tests. While the paper-and-pencil selection technique has many advantages (e.g., more time efficient, objectively scored, favorable psychometric properties, etc.), future research should explore the use of alternative methods to predict teamwork that are more independent of mental abilities, thus helping sort out the relationships between the underlying constructs of general mental abilities and teamwork-related KSAs. These alternative predictor methods might include such techniques as assessment centers, work samples, biodata inventories, or structured interviews.

Another potential limitation is that neither sample was fully advanced in terms of teamwork. Like many organizations, these settings were in transition from traditional work systems to teams. While they were using some degree of teamwork (e.g., for quality and other problem solving tasks), they were not totally developed self-managing work teams. However, it is during these transition periods when staffing teams may be a critical organizational need. Nevertheless,

many teamwork KSAs (e.g., communication, group problem solving, coordination of interdependent work activities, autonomy over work methods, etc.) may still be required to some degree by many jobs, even if they are not truly in a team environment. While this was confirmed for both Studies 1 and 2 via the job analyses, it is likely the case that the teamwork KSAs would be even more important in intensive team environments.

Another limitation is that the research site for Study 2 was less optimal than for Study 1, potentially explaining the relatively weaker effects. Specifically, Study 2 subjects were volunteers in a research study which had no consequence to their jobs. Hence, their test-taking motivation was potentially much lower (Arvey et al., 1990) than for Study 1 subjects. Relatedly, the use of volunteers in Study 2 may have resulted in a sample that did not include employees with inferior teamwork skills, thereby restricting the range of aptitudes measured. Although comparisons of test scores showed no differences in means and variances between studies, those who scored lower in Study 2 may have simply been less motivated rather than less capable. Also, Study 2 had a larger number of supervisors (i.e., 20) each evaluate a smaller number of employees, whereas Study 1 had a smaller number of supervisors (i.e., 5) each evaluate a larger number of employees. Consequently, Study 2 may have allowed for less consistency and standardization in ratings because fewer relative comparisons could be made.

A final limitation to be mentioned is that this research only examined hourly production jobs. While teams are frequently implemented in such settings, future studies should determine how the findings generalize to other types of jobs, industries, country and company cultures, reward systems, and so on. For example, measures reflecting mental abilities, such as the Teamwork Test, may be more useful with more complex and mentally demanding jobs (e.g., white-collar jobs; Hunter & Hunter, 1984). Also, there may be important differences between the various teamwork KSAs for certain types of teams (e.g., executive task force teams, military crews, product development teams, production work teams, quality circles, etc.).

The present research can be extended in several additional ways. For example, future research could examine the outcomes of different decision models for putting together the best mix of team members. One possible strategy is to build heterogeneous teams so as to maximize diversity among team members (Gladstein, 1984; Pearce & Ravlin, 1987), whereas an alternative strategy might be to maximize utility by hiring from the top down on the basis of the same KSAs (Hunter & Hunter, 1984). Future research could help identify which decision strategies are most appropriate under various conditions or team settings.

Future research might also examine whether this research could have potential value for other team management applications besides staffing and selection. For example, the conceptual framework of teamwork KSAs presented in Table 1 could be used as a model around which training and development activities might be developed. Along these lines, the Teamwork Test itself might be useful for diagnosing specific training and development needs of individual team members (i.e., by examining an individual's pattern of scores across items in the various test subscales) and for measuring learning.

Finally, future research also needs to explore the question of which performance measures are most appropriate for evaluating individual team member contributions, and whether such criteria should differ from the perspective of supervisors, peers, or team members themselves. This is suggested in part by the unexpected finding that the Teamwork Test correlated higher with taskwork performance than it did with teamwork performance. One possible explanation for this could be that taskwork performance is more easily and accurately measured than teamwork performance (especially given the relatively longer history of practice refining the measurement of task-related performance). Or perhaps it might be that taskwork performance is more predictable through mental ability, which is what the Teamwork Test largely taps. More research on measuring teamwork versus taskwork performance in the criterion domain is clearly needed.

Conclusion

The domain of individual level KSAs required for teamwork has been previously underdeveloped. With a few exceptions, the various literatures on groups have focused almost exclusively on the management of groups as a collective entity. It should not be forgotten, however, that groups are comprised of individuals who typically are hired one at a time, take home individual paychecks, are concerned with their own personal futures and careers, possess distinct capabilities and interests, and so on. Ultimately, a complete understanding of the entire process of team management will require the integration of all the various levels of analysis, including organizational level issues (e.g., company culture, reward systems, the nature and quality of supervision, etc.), group level processes (e.g., social facilitation, diversity, norms, cohesion, etc.), and individual level attributes, such as those examined in this study. The findings of the present study are, therefore, encouraging since they provide an initial attempt to illustrate how the consideration of individual level teamwork KSAs can have both conceptual and practical value in the management and selection of employees for work teams.

Acknowledgment: Thanks are extended to Chris J. Berger, Stephen G. Green, and F. David Schoorman for their constructive input, and to the research participants (Sarah R. Gaylord and the employees of Weyerhaeuser Company, and Jack Leon and the employees of Inland Container Corporation). An earlier version of this paper was presented at the 1994 Annual Meetings of the Society for Industrial and Organizational Psychology, Nashville, TN.

Appendix

Sample Items from the Teamwork Test

1. Suppose that you find yourself in an argument with several co-workers about who should do a very disagreeable, but routine task. Which of the following would likely be the *most* effective way to resolve this situation?
 - A. Have your supervisor decide, because this would avoid any personal bias.

- B. Arrange for a rotating schedule so everyone shares the chore.
 - C. Let the workers who show up earliest choose on a first-come, first-served basis.
 - D. Randomly assign a person to do the task and don't change it.
2. Your team wants to improve the quality and flow of the conversations among its members. Your team should:
- A. use comments that build upon and connect to what others have already said.
 - B. set up a specific order for everyone to speak and then follow it.
 - C. let team members with more to say determine the direction and topic of conversation.
 - D. do all of the above.
3. Suppose you are presented with the following types of goals. You are asked to pick one for your team to work on. Which would you choose?
- A. An easy goal to ensure the team reaches it, thus creating a feeling of success.
 - B. A goal of average difficulty so the team will be somewhat challenged, but successful without too much effort.
 - C. A difficult and challenging goal that will stretch the team to perform at a very high level, but attainable so that effort will not be seen as futile.
 - D. A very difficult, or even impossible goal so that even if the team falls short, it will at least have a very high target to aim for.

Notes: Correct answers are indicated in italics. From *The Teamwork-KSA Test*, by M. J. Stevens and M. A. Campion, 1993. Copyright 1993 by Authors. Reprinted by permission.

Notes

1. See Stevens and Campion (1994) for a complete review of this topic and literature. Only a brief summary is provided here to establish the background and conceptual basis for this study.
2. We wish to thank an anonymous reviewer for providing this helpful insight.
3. For purposes of future meta-analyses, it should be noted that the correlations between the employment aptitude tests and the criteria in Study 1 were also reported in M. Campion, J. Campion, and Hudson (1994).
4. Details of these analyses are available from the first author.

References

- Aiken, L. S., & West, S. G. 1991. *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Anastasi, A. 1988. *Psychological testing* (6th ed.). New York: Macmillan Publishing Co.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. 1990. Motivational components of test taking. *Personnel Psychology*, 43: 695-716.
- Barrick, M. R., & Mount, M. K. 1991. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44: 1-26.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. 1991. *Technical manual for the Differential Aptitude Tests*. Chicago, IL: Psychological Corporation.
- Borman, W. C. 1991. Job behavior, performance, and effectiveness. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (vol. 2): 271-326. Palo Alto, CA: Consulting Psychologists Press.
- Borman, W. C., & Motowidlo, S. J. 1993. Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations*: 71-98. New York: Jossey-Bass.

- Brief, A. P., & Motowidlo, S. J. 1986. Prosocial organizational behaviors. *Academy of Management Review*, 11: 710–725.
- Campion, M. A., Campion, J. E., & Hudson, J. P. 1994. Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79: 998–1002.
- Cohen, J. 1977. *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cortina, J. M. 1993. What is Coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78: 98–104.
- Dye, D. A., Reck, M., & McDaniel, M. A. 1993. Moderators of the validity of written job knowledge measures. *International Journal of Selection and Assessment*, 1: 153–157.
- File, Q. W., & Remmers, H. H. 1946. Studies in supervisory evaluation. *Journal of Applied Psychology*, 30: 421–425.
- Flanagan, J. C. 1975. *Flanagan Industrial Tests examiner's manual*. Chicago, IL: NCS-London House.
- Flesch, R. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32: 221–233.
- Fry, E. 1968. A readability formula that saves time. *Journal of Reading*, 12: 513–519.
- Gladstein, D. L. 1984. Groups in context: A model of task group effectiveness. *Administrative Sciences Quarterly*, 29: 499–517.
- Goodman, P. S. 1986. *Designing effective work groups*. San Francisco: Jossey-Bass.
- Gordon, R. A. 1968. Issues in multiple regression. *American Journal of Sociology*, 73: 592–616.
- Guzzo, R. A., & Shea, G. P. 1992. Group performance and intergroup relations in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology (vol. 3)*: 269–313. Palo Alto, CA: Consulting Psychologists Press.
- Harris, M. M., & Schaubroeck, J. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41: 43–62.
- Hunter, J. E. 1986. Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29: 340–362.
- Hunter, J. E., & Hunter, R. F. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96: 72–98.
- Jensen, A. R. 1986. g: Artifact or reality? *Journal of Vocational Behavior*, 29: 301–331.
- Kane, J. S., & Lawler, E. E. 1978. Methods of peer assessment. *Psychological Bulletin*, 85: 555–586.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65: 422–427.
- Mayo, E. 1933. *The human problems of an industrial civilization*. New York: Macmillan.
- McDaniel, M. A., Finnegan, E. B., Morgeson, F. P., Campion, M. A., & Braverman, E. P. 1997. *Predicting job performance from common sense*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology. St. Louis, MO.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. 1990. An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75: 640–647.
- Nunnally, J. C. 1978. *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Organ, D. W. 1988. *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Pearce, J. A., & Ravlin, E. C. 1987. The design and activation of self regulating work groups. *Human Relations*, 40: 751–782.
- Ramsay, R. T. 1991. *Manual for the Job Skills Test: Reading*. Pittsburgh, PA: Author.
- Reilly, R. R., & Chao, G. T. 1982. Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35: 1–62.
- Ruch, F. L., & Ruch, W. W. 1980. *Technical report for the Employee Aptitude Survey*. Glendale, CA: Psychological Services, Inc.
- Ruch, W. W., Weiner, J. A., McKillip, R. H., & Dye, D. A. 1985. *Technical manual for the PSI Basic Skills Test*. Glendale, CA: Psychological Services, Inc.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. 1984. Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37: 407–422.
- Schwab, D. P. 1980. Construct validity in organizational behavior research. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior*: 3–43. Greenwich, CT: JAI Press.
- Science Research Associates. 1986. *Reading and Arithmetic Indexes examiner's manual*. Chicago, IL: London House.
- Shore, T. H., Shore L. M., & Thornton, G. C. 1992. Construct validity of self- and peer-evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology*, 77: 42–54.
- Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. 1995. Testing common sense. *American Psychologist*, 50: 912–927.
- Stevens, M. J., & Campion, M. A. 1993. *The Teamwork-KSA Test*. Chicago, IL: NCS-London House.

- Stevens, M. J., & Campion, M. A. 1994. The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20: 503–530.
- Swezey, R. W., & Salas, E. (Eds.). 1992. *Teams: Their training and performance*. Norwood, NJ: Ablex.
- Tett, R. P., Jackson, D. N., & Rothstein, M. 1991. Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44: 703–742.
- Trist, E., & Bamforth, K. 1951. Some social and psychological consequences of the long wall method of coal-getting. *Human Relations*, 1: 3–38.
- Van Scotter, J. R., & Motowidlo, S. J. 1996. Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81: 525–531.
- Walton, R. 1972. How to counter alienation in the plant. *Harvard Business Review*, 50: 70–81.
- Weekley, J. A., & Gier, J. A. 1987. Reliability and validity of the situational interview for a sales position. *Journal of Applied Psychology*, 72: 484–487.
- Wesman, A. G. 1980. *Personnel Tests for Industry: Verbal test manual*. Chicago, IL: Psychological Corporation.
- Wherry, R. J., & Gaylord, R. H. 1943. The concept of test and item reliability in relation to factor pattern. *Psychometrika*, 8: 247–264.