

The Knowledge, Skill, and Ability Requirements for Teamwork: Revisiting the Teamwork-KSA Test's validity

Thomas A. O'Neill*, Richard D. Goffin** and Ian R. Gellatly***

*Department of Psychology, University of Calgary, 500 University Drive, N.W., Calgary, AB, Canada T2N 1N4.
toneill@ucalgary.ca

**Department of Psychology, Social Science Center, University of Western Ontario, London, ON, Canada N6A 5C2

***Department of Strategic Management and Organization, School of Business, University of Alberta, Edmonton, AB, Canada T6G 2R6

The Teamwork – Knowledge, Skills, and Ability (KSA) Test was developed by Stevens and Campion to operationalize their comprehensive taxonomy of teamwork competencies. The test is generally considered 'valid' and has been used frequently in organizations. Our review of the literature found an average criterion validity of .20 for the Teamwork-KSA Test, although there was considerable variability across studies. We could find no research on the item properties, factor structure, or subscale reliabilities, and no extensive investigations of the nomological net of this test. In our field sample, we found subscale reliabilities to be generally inadequate, no meaningful factor structure, and low predictiveness of employees' performance on team-related dimensions. Although the taxonomy it purports to measure is preeminent, the Teamwork-KSA Test itself may have serious limitations.

The practice of organizing employees into teams responsible for critical work activities has become the norm for many different types of organizations (Rapp & Mathieu, 2007; Williams & Allen, 2008). One of the challenges of staffing team-based organizations is selecting applicants who are both technically competent and interpersonally suitable for work in teams. It is the latter of these selection objectives that is most difficult to assess given the divergence of opinions and definitions regarding teamwork processes (for a meta-analytic review of this literature, see LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). A selection test of individual teamwork capabilities that looks particularly promising is the Teamwork Knowledge, Skills, and Abilities (KSA) Test (Stevens & Campion, 1993, 1999). Not only did the Teamwork-KSA Test incorporate a theory-based taxonomy of teamwork capacities (Stevens & Campion, 1994), to our knowledge, this test is a widely known tool for team-based selection. For instance, the prominence of the test in scholarly research is exemplified by the 132 citations of Stevens and Campion's (1994) original article in PsycINFO at the time of this writing.

Although the Teamwork-KSA Test and taxonomy are widely cited in the literature, in actual fact, little is known about the test's reliability and validity. For example, of the studies we found, there were reports of modest overall reliability and there were no reports of subscale reliability coefficients. Moreover, we could find only inconsistent evidence of criterion-related validity (see a review by Allen & West, 2005). To our knowledge, factor analyses, concurrent and discriminant validity, scale-level reliabilities, and item analyses have not been reported even in the test manual (see Stevens & Campion, 1993), despite the fact that this information is integral to informed test use (see Hinkin, 1995). Importantly, we could find no other commercially available, paper-and-pencil test that is designed for measuring the constructs in Stevens and Campion's taxonomy. There is at least one other multiple-choice exam, but it is not publicly available and it was only used in one sample involving military personnel (see Hirschfeld, Jordan, Feild, Giles & Armenakis, 2006). There is a short-answer exam (see Rapp & Mathieu, 2007), but we are not convinced that organizations would use a test requiring trained human assessors when a potentially

valid multiple-choice exam could be scored much faster and less subjectively (i.e., the Teamwork-KSA Test). Behavioral observation scales have been developed (see Taggar & Brown, 2001), but this method requires peer raters who are familiar with the target, which would have no useful application in most selection contexts involving external recruits. Thus, the Teamwork-KSA Test is the logical choice for organizations who wish to measure the constructs in Stevens and Campion's (1994) preeminent taxonomy.

Despite the need for additional psychometric and validity evidence on the Teamwork-KSA Test, it appears that some researchers implicitly consider the test to be 'valid' (e.g., see Hollenbeck, DeRue, & Guzzo, 2004, p. 355). Additionally, a number of test vendors currently market the Teamwork-KSA Test as suitable for personnel decision making in organizations (e.g., Creative Organizational Design, 2009; Ramsay Corp, 2009). We contacted Ramsay Corporation and we were directed to a sales representative from Vangent Inc. who reported that the test is used in dozens of organizations across several industries. From an evidence-based management perspective (e.g., Graen, 2009; Pfeffer & Sutton, 2006; Rousseau, 2006); however, it is disconcerting that researchers and practitioners appear to support the use of the Teamwork-KSA Test while important unanswered questions still exist regarding its psychometric properties and validity.

The impetus for this research was in recognizing that the Teamwork-KSA Test is the only commercially available paper-and-pencil test that aims to measure the constructs in Stevens and Campion's taxonomy, that the test is being used in industry for making decisions about applicants, and that researchers appear to endorse this test (e.g., Hollenbeck et al., 2004). Accordingly, we begin by reviewing the development of this test. We then summarize the results of all the relevant studies pertaining to the Teamwork-KSA Test identified in our computer search of the current literature. Next, we describe the findings of our empirical study bearing on the Teamwork-KSA Test's psychometric properties and construct validity. We report item-total correlations and scale-level reliabilities not published in previous research; we apply factor-analytic methods including confirmatory factor analysis and parallel analysis also not published elsewhere;

and we report new criterion, concurrent, and discriminant validity coefficients in an organization that emphasizes teamwork competencies. The overall purpose of this study was to provide novel insights regarding the validity of the Teamwork-KSA Test.

1. The Teamwork-KSA Test

The development of the Teamwork-KSA test involved a detailed review of the literature on teamwork that focused on team-related KSAs (Stevens & Campion, 1994). The literature review culminated in the hierarchical, taxonomic structure shown in Figure 1. At the apex of the hierarchy is an overall Teamwork-KSA factor, which is then bifurcated into two categories: Interpersonal KSAs and Self-management KSAs. The Interpersonal category comprises the dimensions of Conflict Resolution, Collaborative Problem Solving, and Communication KSAs, whereas the Self-management category comprises the dimensions of Goal Setting and Performance Management, and Planning and Coordination KSAs (hereafter, the five lower-level dimensions are referred to as scales). Figure 1 shows the three hierarchical levels at which the Teamwork-KSA Test provides scores. However, only the overall score is recommended for selection purposes (Stevens & Campion, 1993, 1999). Interpersonal and Self-management category and scale scores were suggested for use in employee development and training contexts, possibly because personnel selection can be highly litigious and, accordingly, only the most reliable scores should be used for that purpose. Nonetheless, employee development and training are also highly consequential activities from the perspective of both employees and organizations, and, as such, one could argue that only highly reliable scale scores should be used for those purposes too.

1.1. Test construction

Stevens and Campion (1993, 1999) began their development of the Teamwork-KSA Test by writing 46 situational items, each of which described a hypothetical teamwork scenario. Four multiple-choice response alternatives accompanied each item, and based on their literature

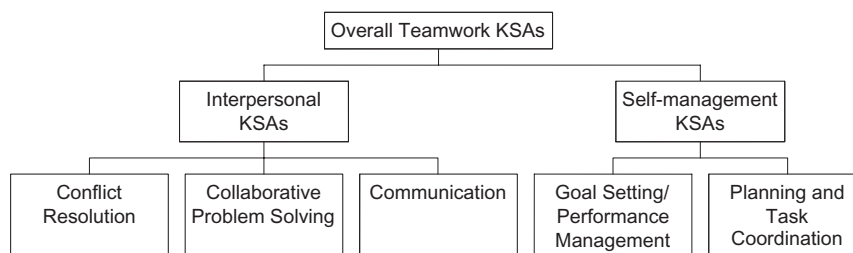


Figure 1. Hierarchical levels at which the Teamwork-Knowledge, Skills, and Abilities (KSA) Test can be scored.

review, only one was identified as the best approach to resolving the problem presented. The task for the respondent is to indicate which of the four multiple-choice alternatives represents the most effective course of action, and the responses are then coded 0 (*incorrect*) or 1 (*correct*). After developing this initial test, 234 undergraduates responded to the items. It is unclear whether these students had teamwork experience. Based on the psychometric properties of the 46 test items, the authors retained the strongest 35 items for inclusion in the final Teamwork-KSA Test.

Stevens and Campion's (1999) report does not make it clear that items were written with a specific construct in mind. Although the teamwork taxonomy was used to guide the writing of items, the authors did not appear to have directly targeted a specific construct within their framework with each item. In order to determine the scale to which each item would be assigned, Stevens and Campion conducted a small sample content validity study. Four judges assigned items to scales, and it was these assignments that determined the items that would comprise each scale. Findings indicated that judges categorized items under each of the five scales relatively consistently, although the final number of questions in each scale varied: Conflict Resolution (4); Collaborative Problem Solving (8); Communication (12); Goal Setting and Performance Management (5); and Planning and Task Coordination (6).

Before reviewing the research findings on the Teamwork-KSA Test, a cursory overview of established test development practices, which has implications for Stevens and Campion's (1993, 1999) approach to developing the Teamwork-KSA Test, is in order. We rely on practices described in Aguinis, Henle, and Ostroff (2001), Hinkin (1995), Murphy and Davidshofer (2005), and Nunnally (1978). First, the attribute or construct to be measured should be clearly defined. The meaning of the overall Teamwork-KSA Test score, however, does not appear to correspond to a defined construct in Stevens and Campion's (1999) taxonomy. This is concerning because, as previously mentioned, the overall score is the most reliable and most recommended for research and practice. Second, it is advisable to write items deductively; that is, with the construct guiding the content of the items. Stevens and Campion's (1999) description of the Teamwork-KSA Test construction leads the reader to assume that items were written somewhat inductively. Judges determined to which construct items belonged at the final stage of the test construction instead of earlier. Third, factor analysis should be used to confirm the test's structural validity; that is, the extent to which the empirical structure of the test conforms to its theoretical structure (Hinkin, 1995; Loevinger, 1967). Unfortunately, we could find no published accounts investigating the factor structure of this test. Thus, the appropriateness of interpretations at various levels of the test's scoring hierarchy (Figure 1) needs further research. Fourth, Stevens

and Campion (1999) conducted two criterion-validity studies and found mixed, but general support, for the association between the Teamwork-KSA Test scores and other ratings of job performance. However, the sample sizes in these studies were small for determining item functioning and for long-term validation purposes ($N = 70$ and 72).

Taken together, the development of the Teamwork-KSA Test seems to have deviated from some established test construction principles. We make this point only because of the prevalence of the test and because the rigor of the test construction methodology is integral to any test's validity (see Guion, 1998; Jackson, 1999; Kline, 1984). We next review the studies wherein the reliability and criterion validity of the Teamwork-KSA Test has been reported.

1.2. Reliability of the Teamwork-KSA Test

We found six samples that reported Cronbach's alpha for the overall Teamwork-KSA Test score: McClough and Rogelberg (2003), $\alpha = .59$; Leach, Wall, Rogelberg, and Jackson (2005), $\alpha = .70$; Chen, Donahue, and Klimoski (2004), Time 1 $\alpha = .64$ and Time 2 $\alpha = .82$; Martín-Pérez, Martín-Cruz, Pérez-Santana, Hernangómez-Barahona, and Martín-Sierra (2010), Time 1 $\alpha = .68$ and Time 2 $\alpha = .75$; and Stevens & Campion, (1999), after deleting two unidentified items with negative item-total correlations, $\alpha = .80$ and $.81$ in two samples. It is noteworthy that, despite deleting two items during the reliability analysis, it appears that those items were retained for the published test version. Finally, we found two test-retest reliability coefficients of $.60$ and $.72$ reported by Chen et al. and Martín-Pérez et al., respectively. Time intervals appeared to be approximately 15 weeks.

It is unclear exactly how high test reliability coefficients should be (Bobko, 2001), but reliability should be as high as possible when making important personnel decisions. Reliability affects the precision of measurement and therefore the confidence one can have regarding whether two test scores are reliably different. Lower levels of reliability may be tolerable when decisions do not require precision, although most uses of test scores, including development applications, would probably benefit from the highest possible level of reliability. Overall, the range of reliabilities reported in the extant literature suggests the overall Teamwork-KSA Test score has moderate reliability. Category- and scale-level reliabilities are still open questions; however, as it appears, none have been published, and Leach et al. (2005) found that those reliabilities were too low to report. Although Stevens and Campion (1993) advised against using those lower-level scores for selection, Stevens and Campion suggested that they could be used for development and training needs. Thus, category- and scale-level scores have the potential to affect individuals' treatment by an organization.

31.3. Criterion validity of the Teamwork-KSA Test

Our literature search identified eight articles describing nine samples where criterion validity coefficients for the Teamwork-KSA Test were reported (see Table 1). Samples were composed of ad hoc undergraduate laboratory teams (2), undergraduate student project teams (2), and manufacturing and production teams (5). Five studies were conducted at the individual level, three were conducted at the team level, and one was conducted at both the individual and the team level. In team-level studies, the mean of team members' scores on the Teamwork-KSA Test was correlated with overall team performance scores. Criterion variables at the individual level included supervisor, peer, and subordinate ratings, and at the team level, team project performance, satisfaction, safety, and so forth. Although it would be interesting to calculate average sample-weighted mean effects sizes, corrected for artifacts, studies varied too much by level of analysis and criteria to be meta-analytically summarized (see Osburn, Callender, Greener & Ashworth, 1983; Oswald & Johnson, 1998).

Table 1 lists the nine studies from our literature search that included at least one criterion-validity coefficient. Although it is premature to conduct a meta-analysis, of the 33 (non-combined sample) coefficients in Table 1,¹ it is instructive to note that the average criterion validity is .20, suggesting that the average predictiveness of the Teamwork-KSA Test is neither large nor small. The standard deviation (*SD*) of these coefficients, however, is 0.21, which indicates substantial variation in criterion correlations across criteria and studies. This is probably in part due to the range of correlations, which was from $-.30$ (Kottke, 2008) to $.56$ (Stevens & Campion, 1999). Average prediction at the team level was $.25$ ($SD = 0.17$, 10 validity coefficients), and the average prediction at the individual level was $.17$ ($SD = 0.22$, 23 validity coefficients). It is somewhat surprising that the validity coefficients appear slightly larger at the team than at the individual level because the test was designed to predict individual-level outcomes (Stevens & Campion, 1999). Overall, we interpreted these results as providing mixed support for the criterion validity of the Teamwork-KSA Test, which is consistent with the conclusion reached in an earlier review (Allen & West, 2005).

1.4. Criterion validity and teamwork versus taskwork

The difference between taskwork and teamwork has been pointed out by several authors who suggested that both are needed to perform effectively in a team-based environment, but that a different skill set underlies each (Marks, Mathieu, & Zaccaro, 2001; Rapp & Mathieu, 2007; West & Allen, 1997). Teamwork variables are team process activities – cognitive, verbal, and behavioral – that

are used to organize the team's work in order to meet performance targets (Marks et al., 2001). Taskwork involves the technical and job-specific aspects of the work itself (Stevens & Campion, 1999). It was argued by Stevens and Campion (1999) that the Teamwork-KSA Test should correlate higher with teamwork criteria than with taskwork criteria because Stevens and Campion's (1994, 1999) taxonomy is directed specifically at teamwork variables. Nonetheless, three of three samples that compared teamwork criteria to taskwork criteria using supervisor ratings found that the Teamwork-KSA Test was more strongly related to taskwork than it was to teamwork (Morgeson, Reider, & Campion, 2005; Stevens & Campion, 1999). In one of those samples, there was an exception, wherein the Teamwork-KSA Test was more strongly correlated with peer ratings of teamwork than it was with taskwork (Stevens & Campion, 1999, Sample 2).

1.5. The present study

Our review of the research suggests that important questions remain regarding the construct validity and psychometric properties of the Teamwork-KSA Test, specifically, there is an immediate need to examine:

1. The extent to which the Teamwork-KSA Test reliably measures a single construct and several narrower constructs.
2. The extent to which Teamwork-KSA Test scores at various levels of the taxonomy (see Figure 1) can be meaningfully interpreted.
3. The extent to which the Teamwork-KSA Test correlates well with variables to which it is theoretically related (concurrent validity), and the extent to which it correlates minimally with theoretically unrelated variables (discriminant validity).
4. The criterion validity of the Teamwork-KSA Test in team-based organizations.

In order to investigate each of these four issues, we began by evaluating the psychometric properties of item, scale, category, and overall Teamwork-KSA Test scores (item 1 as described earlier). Following item and reliability analyses, we applied a variety of factor-analytic techniques to shed light on the Teamwork-KSA Test's structural validity (as described earlier; see Loevinger, 1967, item 2). Next, we investigated the concurrent and discriminant validity of the Teamwork-KSA Test (item 3 as described earlier). Our focus here was on correlations with personality and test-taking states (e.g., test-taking motivation) instead of general mental ability (GMA) because the latter has already been shown to be highly related to Teamwork-KSA Test scores (with corrected correlations in the 0.90s, see Stevens & Campion, 1999). Thus, instead of measuring GMA, we used our testing time to measure other constructs relevant to an evaluation of the Teamwork-KSA Test (e.g., personality). Finally, we

Table 1. Summary of research on the criterion validity of the Teamwork-Knowledge, Skills, and Abilities Test

Publication	Sample	Individual or team-level study?	Sample size	Criterion	Criterion validity coefficient	Effect size ^a								
Stevens and Campion (1999)	1. Pulp mill	Individual	70	Supervisor teamwork rating	.44*	Medium								
				Supervisor taskwork rating	.56*	Large								
				Supervisor overall performance rating	.52*	Large								
				Supervisor teamwork rating	.21*	Small								
				Supervisor taskwork rating	.25*	Small								
				Supervisor overall rating	.23*	Small								
				Self teamwork rating	-.12	Small ^b								
				Self taskwork rating	.07	Negligible								
				Self overall rating	-.06	Negligible ^b								
				Peer teamwork rating	.23*	Small								
Stevens and Campion (1999)	2. Box manufacturing	Individual	72	Peer taskwork rating	-.04	Negligible ^b								
				Peer overall rating	.21*	Small								
				Supervisor teamwork rating	.32*	Medium								
				Supervisor taskwork rating	.40*	Medium								
				Supervisor overall rating	.37*	Medium								
				Team satisfaction	.10	Small								
				Project team grade	.20	Small								
				Self ratings of job performance	.06	Small								
				Peer ratings of job performance	-.02	Negligible ^b								
				Overall ratings of job performance	.27*	Small								
Miller (2001)	3. Undergraduate project teams	Team	42	Team effectiveness	.25	Small								
				Team safety	.17	Small								
				Team process	-.04	Small ^b								
				Team role structure	.24	Small								
				Observer ratings of team performance	.31*	Medium								
				Peer ratings of team performance	.34*	Medium								
				Self-efficacy for teamwork	.19*	Small								
				Observer ratings of planning and task coordination	.36*	Medium								
				Observer ratings of collaborative problem solving	.20	Small								
				Observer ratings of communication skills	.49*	Medium								
Stevens, Jones, and Fisher (2002)	4. Large metal refinery	Individual	458	Supervisor ratings of contextual performance	.32*	Medium								
				Supervisor ratings of task performance	.36*	Medium								
				Managerial ratings of team performance	.51*	Large								
				Team orientation	-.30*	Medium ^b								
				Team satisfaction	.10	Small								
				Perceptions of team success	-.26*	Medium ^b								
				McClough and Rogelberg (2003)	5. Undergraduate laboratory teams	Individual	162-170	Observer ratings of collaborative problem solving	.20	Small				
								Observer ratings of communication skills	.49*	Medium				
								Supervisor ratings of contextual performance	.32*	Medium				
								Supervisor ratings of task performance	.36*	Medium				
Managerial ratings of team performance	.51*	Large												
Ellis, Bell, Ployhart, Hollenbeck, & Ilgen (2005)	6. Undergraduate laboratory teams	Team	65					Team orientation	-.30*	Medium ^b				
								Team satisfaction	.10	Small				
								Perceptions of team success	-.26*	Medium ^b				
								Morgeson et al. (2005)	7. Steel-processing teams	Individual	90	Team orientation	-.30*	Medium ^b
												Team satisfaction	.10	Small
				Perceptions of team success	-.26*	Medium ^b								
				Leach, Wall, Rogelberg, & Jackson (2005)	8. Steel and photographic paper production teams	Team	41					Team orientation	-.30*	Medium ^b
												Team satisfaction	.10	Small
												Perceptions of team success	-.26*	Medium ^b
												Kottke (2008)	9. Undergraduate project teams	Individual
Team satisfaction	.10	Small												
Perceptions of team success	-.26*	Medium ^b												

Notes: * $p < .05$. All criterion variables were keyed such that greater scores were indicative of higher performance. ^aAbsolute effect size based on Cohen (1992): negligible ($r < .10$); small ($r = .10-.29$); medium ($r = .30-.49$); and large ($r > .50$). ^bEffect is in the direction opposite to prediction.

presented new criterion-related validity evidence on the Teamwork-KSA Test at its three hierarchical levels (i.e., overall, category, and scale scores) in an organization that is designed around teamwork (item 4 as described earlier). Consistent with the unitary perspective on validity (Guion, 1998; Messick, 1991), our overarching goal was to provide new evidence regarding the extent to which common inferences based on Teamwork-KSA Test scores are justifiable and appropriate, and, especially, the extent to which test score inferences seem likely to serve the test's stated purposes of selection, development, and training (Stevens & Campion, 1993, 1994, 1999).

2. Method

2.1. Participants and procedure

Participating in this study were 268 job candidates (74% male) who were within a selection process for jobs in a large industrial complex. The core operations of this business involved the mining of oil-sand deposits, extraction of bitumen (a sticky, tar-like substance), and the upgrading of the raw bitumen into synthetic oil. All candidates were required to take safety tests, receive instructions from superiors, and demonstrate good standing for the duration of a probationary period before they were considered permanent employees. During their instruction and training on safety modules, which included the safety examinations that had to be passed, candidates were asked to voluntarily complete the Teamwork-KSA Test along with the other scales, some of which were not included in this study. Applicants had to demonstrate successful completion of high school, which suggests their reading capabilities easily surpassed the eighth-grade reading level at which the Teamwork-KSA Test was written (see Ramsay Corp, 2009). To our knowledge, all job candidates participated.

The present study follows a typical predictive validity design. Predictive instruments, such as the Teamwork-KSA Test, were administered during the hiring process and the criterion, supervisor ratings of demonstrated teamwork behaviors, were collected 6–12 months later. Predictive designs, relative to concurrent designs, are advantageous because stronger inferences about the forecasting of job performance can be drawn (see Guion, 1998; Murphy & Davidshofer, 2005). Moreover, that participants were in an applicant mindset is consistent with Stevens and Campion's (1999) findings that were stronger for applicant than for incumbent samples, likely because the former are more motivated to perform well on the test.

2.2. Job and teamwork analysis

Prior to the study proper, a multi-method job analysis was conducted to determine the competencies needed to

perform effectively in this organization. Subject-matter (job) experts were drawn from all core operations including the specific jobs being hired for. In total, the job analysis involved 43 people who completed the Common Metric Questionnaire (CMQ; Harvey, 1993) and 51 people who were interviewed using the critical incident technique. The results of this analysis revealed the importance of three team-related competencies (communication, cooperation, and teamwork) that were critical to all jobs. The entire organization was structured around teams, and its continuous-process technology required that highly interdependent teams in the mining, extraction, and upgrading operations, as well as the teams that support these operations, coordinate seamlessly.

2.3. Focal measures

2.3.1. Teamwork-KSA Test

Stevens and Campion's (1993) 35-item multiple-choice Teamwork-KSA Test was administered. The psychometric properties of this scale were discussed earlier.

2.3.2. Job performance

Job performance ratings from individuals who satisfied all of the following criteria were utilized: (a) observed the new hire on the job for at least 6 months; (b) possessed administrative authority to direct the worker in his or her job (e.g., managers, team leaders, trainers); (c) were qualified in all areas of the job in which the person worked; and (d) rated themselves as being at least 'fairly familiar' with the worker. Performance ratings were obtained from two raters and averaged together whenever possible, which occurred in 42% of the cases (i.e., for 113 targets).

We used the relative percentile method (RPM) to structure the performance rating scales. The accuracy and validity of the RPM has been demonstrated in several studies (e.g., Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996; Goffin, Jelley, Powell, & Johnston, 2009; Wagner & Goffin, 1997). This approach resulted in a rating within the 0–100 range for each of three teamwork-related performance dimensions identified by the job analysis described previously. RPM ratings of communication, cooperation, and teamwork were aggregated to form a unit-weighted composite index of team-relevant performance in this organization.

2.4. Additional measures

In addition to the focal study measures described earlier, several scales were administered in order to investigate the concurrent and discriminant validity of the Teamwork-KSA Test. These measures were administered at the same time as the Teamwork-KSA Test.

2.4.1. Personality

The Six-Factor Personality Questionnaire (SFPQ) was used to assess the 'Big Five' personality variables included

in this study. The SFPQ is similar to other big five measures except that it bifurcates conscientiousness into achievement and methodicalness. The validity and psychometric properties of that measure have been supported in numerous studies (see Jackson, Ashton, & Tomes, 1996; Jackson, Paunonen, Fraboni, & Goffin, 1996; Jackson, Paunonen, & Tremblay, 2000). As was the case for all measures except job performance and the Teamwork-KSA Test, responses were made on a standard 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

2.4.2. Social desirability

Paulhus's (1991) Balanced Inventory of Desirable Responding was used to measure impression management and self-deceptive enhancement. One item from each scale was deleted because they were potentially objectionable to some candidates, which left 19 items per scale. We followed the traditional dichotomous scoring method recommended by Paulhus (see also McFarland & Ryan, 2006); thus, the impression management and self-deceptive enhancement scores both ranged from 0 from 19.

2.4.3. Test-taking motivation

Arvey, Strickland, Drauden, and Martin's (1990) 10-item scale was used to measure test-taking motivation. The reliability and validity of that scale has been supported (see Arvey et al., 1990; Schmit & Ryan, 1992; O'Neill, Goffin, & Gellatly, 2010).

2.4.4. Test-taking anxiety

The Test Attitude Inventory (Spielberger, 1980) was used to assess the worry and emotionality components of test-taking anxiety (10 items each). The inventory has been used in several studies in which its reliability and validity has been supported (e.g., McCarthy & Goffin, 2004, 2005; Spielberger & Vagg, 1995).

3. Results

3.1. Properties of Teamwork-KSA Test items

Where multiple-choice, dichotomously scored items like the Teamwork-KSA Tests are concerned, item means, also known as item difficulties or *p* values, in the neighborhood of .50 are desirable (Nunnally, 1978). Items with means close to 0 or 1 provide little discrimination among candidates. It is also generally preferred to have every item's mean close to 0.50, and an average item mean of 0.50, instead of having a large variance in item means (Murphy & Davidshofer, 2005). Table 2 shows that the item means for the Teamwork-KSA Test ranged from 0.30 to 0.93, and that the grand (test) mean was 0.67 (*SD* = 0.16). Overall, a reasonable proportion of item

difficulties were near their desired 0.50 target (cf. Allen & Yen, 1979), although a few items provide little discrimination (e.g., Conflict Resolution item 3, Planning and Task Coordination item 3).

Turning to item-total correlations, Table 2 presents each of the Teamwork-KSA items' correlation with their respective scale (e.g., Conflict Resolution), category (e.g., Interpersonal KSA), and the overall Teamwork-KSA score, corrected for part-whole overlap (Furr & Bacharach, 2008). Item-total correlations provide an indication of the relation between the test item and the construct to which it is presumed to belong (Murphy & Davidshofer, 2005). Item-total correlations should always be positive and at least moderate in magnitude (Guion, 1998; Murphy & Davidshofer, 2005). It is important that item-total correlations are positive because such values indicate that the item correctly discriminates people who do well from people who do poorly on the scale of interest. More generally, it is an indication of the extent to which the item measures the same construct as the rest of the test or scale (Allen & Yen, 1979). Common test development practices are to check that items with negative item-totals were miskeyed, and if not, to delete these items (Nunnally, 1978).

The item-total correlations for the Teamwork-KSA overall test score, shown in Table 2, had a range of $-.09$ to $.43$ ($M = 0.21$), and two values were negative (see also Stevens & Campion, 1999). The range of item-totals for the categories of Self-management and Interpersonal KSA was $-.06$ to $.39$ ($M = 0.20$), and the range of item-totals for the scales was $-.04$ to $.36$ ($M = 0.15$). The mean item-totals by scale were as follows: Conflict Resolution ($-.02$), Communication (.12), Collaboration and Problem Solving (.12), Goal Setting and Performance Management (.23), and Planning and Task Coordination (.19). Nunnally (1978) suggested that item-total correlations above .30 are 'usually considered good' (p. 263; see also Aguinis et al., 2001, p. 32). We recognize that this is a rule-of-thumb and is not an absolute cutoff for any item's inclusion in a scale. Certainly, very long tests may still be reliable when item-totals are lower. Nevertheless, most item-totals on the Teamwork-KSA Test were below .30, and a few items were operating in opposition to the overall test score (negative item-totals). Specifically, there were two negative item-totals at the category level and five at the scale level. Average interitem correlations tended to be less than .10 and two were negative. Perhaps of greatest concern is in regards to the Conflict Resolution scale, which had an average item-total correlation of $-.02$ with its own scale.

3.2. Properties of Teamwork-KSA Test

Table 3 describes psychometric properties of the Teamwork-KSA Test overall, category and scale scores. The range of scores on the overall Teamwork-KSA Test

Table 2. Properties of the Teamwork-Knowledge, Skills, and Abilities Test items

Item	<i>M</i> (<i>p</i>)	<i>SD</i>	<i>r</i> _{is}	<i>r</i> _{ic}	<i>r</i> _{it}	Average interitem correlation	Average -item-criterion correlation
Conflict resolution 1	.79	.41	-.04	.27	.39	.10	.06
Conflict resolution 2	.54	.50	-.04	-.01	-.03	-.01	.02
Conflict resolution 3	.93	.25	.00	.21	.20	.06	-.10
Conflict resolution 4	.74	.44	-.01	.06	.03	.01	-.02
Conflict resolution mean	.75	.40	-.02	.13	.15	.04	-.01
Conflict resolution <i>SD</i>	.16	.11	.02	.13	.19	.05	.07
Communication 1	.42	.49	.10	.17	.23	.06	.13
Communication 2	.81	.39	.12	.20	.22	.07	.01
Communication 3	.54	.50	.16	.29	.33	.10	.03
Communication 4	.87	.33	.10	.11	.11	.03	.02
Communication 5	.60	.49	.23	.21	.24	.07	-.06
Communication 6	.88	.32	.01	.07	.09	.03	.04
Communication 7	.74	.44	.15	.08	.11	.04	-.06
Communication 8	.66	.47	.22	.28	.30	.10	-.02
Communication 9	.39	.49	-.02	.04	.03	.01	-.11
Communication 10	.62	.49	-.03	-.06	-.09	-.03	.05
Communication 11	.62	.49	.24	.37	.35	.10	.06
Communication 12	.56	.50	.18	.16	.17	.06	-.01
Communication mean	.64	.45	.12	.16	.17	.05	.01
Communication <i>SD</i>	.16	.66	.09	.12	.13	.04	.07
Collaborative problem solving 1	.44	.50	.10	.11	.10	.03	.01
Collaborative problem solving 2	.57	.50	.14	.19	.20	.07	.01
Collaborative problem solving 3	.89	.31	.11	.04	.07	.02	.01
Collaborative problem solving 4	.53	.50	.28	.35	.40	.12	-.01
Collaborative problem solving 5	.54	.50	.31	.34	.38	.11	-.03
Collaborative problem solving 6	.65	.48	.04	.10	.09	.03	.07
Collaborative problem solving 7	.30	.46	.11	.15	.17	.05	-.03
Collaborative problem solving 9	.48	.50	.36	.39	.43	.13	.02
Collaborative problem solving mean	.55	.47	.19	.22	.24	.08	.01
Collaborative problem solving <i>SD</i>	.16	.06	.12	.13	.14	.02	.07
Goal setting & performance management 1	.69	.47	.25	.24	.18	.05	.01
Goal setting & performance management 2	.82	.39	.15	.16	.17	.05	.00
Goal setting & performance management 3	.83	.37	.22	.22	.23	.07	.03
Goal setting & performance management 4	.63	.48	.20	.34	.36	.11	-.03
Goal setting & performance management 5	.88	.32	.31	.34	.31	.09	-.09
Goal setting & performance management Mean	.77	.41	.23	.26	.25	.08	-.02
Goal setting & performance management <i>SD</i>	.11	.07	.06	.08	.08	.02	.05
Planning and task coordination 1	.87	.34	.04	.07	.18	.06	.08
Planning and task coordination 2	.70	.46	.24	.29	.28	.08	.06
Planning and task coordination 3	.90	.30	.26	.29	.26	.08	.00
Planning and task coordination 4	.79	.41	.16	.20	.28	.08	.09
Planning and task coordination 5	.69	.46	.24	.32	.40	.12	-.01
Planning and task coordination mean	.79	.39	.19	.23	.28	.08	.04
Planning and task coordination <i>SD</i>	.10	.07	.09	.10	.08	.02	.04
Grand mean	.67	.44	.15	.20	.21	.06	.01
Grand <i>SD</i>	.16	.07	.11	.12	.13	.04	.05

Notes: *N* = 268 except for criterion correlations, *N* = 225. *M* (*p*) = item mean, or the probability an individual scored the item correct; *SD* = item standard deviation; *r*_{is} = corrected item-scale correlation; *r*_{ic} = corrected item-category correlation; *r*_{it} = corrected item-Teamwork-Knowledge, Skills, and Abilities Test correlation. Values in boldface are negative item-total correlations, negative average interitem correlations, or negative item-criterion correlations.

was from 7 (20% correct) to 33 (94% correct). Means (difficulties) and *SD*s were generally acceptable. Turning to Cronbach's alpha, our findings suggest that coefficients were not high in some cases. Consider guidelines advanced by Murphy and Davidshofer (2005) who suggested that reliabilities not fall lower than .60. When the test is used for important decision making, values below .90 are less than ideal (Nunnally & Burnstein, 1994). We

realize, however, that tests used in personnel selection have reliabilities that rarely exceed .90. Some commonly used selection tests have average meta-analytic internal consistency reliabilities of .81 (integrity; Ones, Viswesvaran, & Schmidt, 2003), .83 (GMA; Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003); and .76 (average of the 'Big Five' personality factors; Viswesvaran & Ones, 2000). An inspection of Table 3 indicates that the

Table 3. Properties of the Teamwork-KSA Test, categories, and overall scores

Teamwork-KSA test score	Number of items	M	SD	α	Range of interitem correlations
Interpersonal KSAs	25	15.76	3.40	.58	-.13 to .35
Conflict resolution	4	3.00	.80	.00	-.06 to .08
Communication	12	7.53	1.92	.36	-.12 to .19
Collaborative problem solving	9	5.23	3.36	.46	-.06 to .33
Self-management KSAs	10	7.79	1.81	.56	-.15 to .25
Goal setting and performance management	5	3.84	1.13	.42	.07 to .25
Planning and task coordination	5	3.94	1.14	.37	-.15 to .21
Teamwork-KSA overall	35	23.54	4.60	.71	-.16 to .35

Notes: $N = 268$. KSA = Knowledge, Skills, and Abilities; M = aggregate mean score; SD = aggregate standard deviation; α = standardized Cronbach's alpha.

reliabilities for the Teamwork KSA Test are, not surprisingly, highest for the overall test score ($\alpha = .71$). The average scale-score alpha was .32 and the average category-score alpha was .57. These scale- and category-score reliabilities did not reach those of commonly used tests, cited earlier, nor did they reach general rule-of-thumb recommendations (e.g., Murphy & Davidshofer). The alpha of .00 for the Conflict Resolution was somewhat unexpected, as this indicates that the Conflict Resolution items have no variance in common. More generally, unreliability increases standard errors of measurement, which contributes to imprecision and reductions in test utility in personnel decision-making practices.

3.3. Confirmatory factor analyses (CFA)

In order to investigate the factor structure of the Teamwork-KSA Test, CFAs on the item covariance matrix were conducted using *Equations* (EQS; Bentler, 2003). Based on the development and intended use of the Teamwork-KSA Test, the fit of three models was assessed: (a) a one-factor model representing the Teamwork-KSA Test overall score; (b) a two-factor model representing the categories of Interpersonal and Self-management KSAs; and (c) a five-factor model representing each of the five scales described earlier. For the two- and five-factor models, we allowed the factors to correlate because the respective constructs appeared not to be orthogonal to each other. Maximum likelihood estimation (robust method) was used to estimate model parameters, and three fit indices were reported to determine the acceptability of the model: the relative noncentrality index (RNI; McDonald & Marsh, 1990), the standardized root mean residual (SRMR), and the root mean square error of approximation (RMSEA). The comparative fit index (CFI) was recommended by Bentler (1990), but we report instead the RNI because it yields identical values to the CFI with the exception that the former is not truncated at 1.0 (i.e., the RNI can properly exceed 1.0, the CFI cannot; Goffin, 1993). SRMR and RMSEA were recommended by Hu and Bentler (1999). Values of RNI above

Table 4. Confirmatory factor analysis goodness-of-fit indicators for the Teamwork-Knowledge, Skills, and Abilities Test

Model	Satorra-Bentler χ^2	df	RNI	SRMR	RMSEA
Single factor	725.74***	560	.72	.06	.03
Two factor	725.36***	559	.72	.06	.03
Five factor	710.82***	550	.73	.06	.03

Notes: *** $p < .001$. $N = 268$. df = degrees of freedom; RNI = relative noncentrality index; SRMR = standardized root mean residual; RMSEA = root mean square error of approximation.

0.95 and values of SRMR and RMSEA below .08 and .06, respectively, suggest acceptable model fit (Hu & Bentler, 1999). In addition to these three fit statistics, we report the Satorra-Bentler scaled chi-squared index (χ^2).

The display in Table 4 indicates that, although the SRMR and RMSEA values were in the acceptable range, the RNI was far too low to suggest any of the models provided an acceptable fit to the data. RNI essentially compares the difference between the model-implied covariance matrix and the independence model (zero covariance) covariance matrix. When RNI is low the model-implied covariance matrix cannot be accepted as a better fit than the fit of the independence model, which assumes no correlation among items (see Kline, 2004). Thus, we rejected these three models. This is not inconsistent with the findings regarding the Teamwork-KSA Test's reliability, presented earlier, which suggested that the items tended not to hold together well.

3.4. Exploratory factor analyses

Because none of the *a priori* CFA models were found to fit acceptably, we conducted exploratory factor analyses (EFA) to further investigate the underlying factor structure of the Teamwork-KSA Test. Our tactic was to first run an EFA using principal axis factoring (PAF) with oblimin rotation, thereby allowing for correlated factors to gain an overall understanding of how many factors may underlie the Teamwork-KSA Test. The scree plot suggested that the percentage of common variance among the items tended to be low. Specifically, one factor was

substantially larger than the others, but that first factor was very small, explaining only 9.20% of the interitem covariance. The second, third, and fourth factors accounted for an additional 3.13%, 2.96%, and 2.20% of the covariance, respectively. Accordingly, far too many factors would be needed to explain a reasonable proportion of the Teamwork-KSA Test items' covariance.

Arguably, the most accurate and least subjective approach to determining the number of factors underlying a data set is parallel analysis (for an overview, see Hayton, Allen, & Scarpello, 2004; for its accuracy, see Velicer, Eaton, & Fava, 2000). The logic of parallel analysis is that the number of meaningful factors in a data set should be the number of factors observed with sizes significantly greater than corresponding factor sizes found in randomly generated parallel data sets. Using syntax available from O'Connor (2000), we generated 1,000 random parallel data sets and then computed the 95% confidence interval around the mean factor sizes extracted from each data set. The number of factors in our data set exceeding that cutoff was six; accordingly, six factors were retained from the parallel analysis.

We forced a six-factor solution using PAF and attempted to interpret the resulting factors using oblimin rotation. This solution accounted for only 26.33% of the variation in the 35 Teamwork-KSA Test items, meaning that these six factors were not sufficiently explaining the covariation in Teamwork-KSA Test items. This is consistent with the somewhat low Cronbach's alpha reliabilities and the CFA results reported earlier, all of which point to a lack of substantial common variance underlying the Teamwork-KSA Test items. That stated, we pressed on and attempted to interpret the factors at various levels of Delta in the Statistical Package for the Social Sciences (SPSS; SPSS Inc., Chicago, IL; 0.5, -0.5, -2.0, -4.0), which allows differing levels of covariance among the factors. None of these rotations produced an interpretable factor structure, and none of the items clustered in ways consistent with the theory that inspired this test (i.e., items belonging to the same scale always loaded highest on several different factors; simple structure was lacking). Overall, the EFAs and CFAs were consistent in casting doubt on the structural validity of the Teamwork-KSA Test.

3.5. Concurrent validity

Concurrent validity of the Teamwork-KSA Test would be supported if the test correlated, in the predicted direction, with constructs to which it is hypothetically related (Cronbach & Meehl, 1978). Appropriately chosen personality traits represent a class of individual difference variables that should be positively related to the Teamwork-KSA Test. Considering that the Teamwork-KSA Test aims to measure Communication, Collaborative Problem Solving, and so forth, it is difficult to imagine that personality traits would be independent of Teamwork-KSA Test

scores. For example, studies have found that personality predicts individuals' performance in interpersonal (Mount, Barrick, & Stewart, 1998) and team settings (Stewart, Fulmer, & Barrick, 2005), which is the same criterion that the Teamwork-KSA Test intends to measure (i.e., an individual's likelihood of succeeding in a teamwork setting). Moreover, other ability-based tests have shown to be correlated with personality when the underlying constructs measured are theoretically related to personality (emotional intelligence; Joseph & Newman, 2010; Schulte, Ree, & Carretta, 2004). Accordingly, unreasonably low correlations between the appropriate personality constructs and the construct of overall Teamwork KSA are theoretically unlikely.

In order to assess the concurrent validity of the Teamwork-KSA Test with personality, we asked a small sample ($N=6$) of subject matter-experts (e.g., faculty members and graduate students of industrial and organizational psychology) to rate the relevance of each of several personality traits for predicting Teamwork-KSA Test scores. The raters were provided with the definition of each trait and the content domain of the Teamwork-KSA Test. Regarding the latter, we defined the Teamwork-KSA Test score according to its scale definitions (Stevens & Campion, 1994), and the statement that the Teamwork-KSA Test measures 'the extent to which an individual has the capacity (i.e., the Knowledge, Skills, and Abilities) to perform effectively in a team setting.' The *a priori* relevance ratings were provided on a 5-point Likert scale [$-2 = a$ moderate or stronger negative relation ($r < -.30$); $-1 = slight$ negative relation ($r = -.10$ to $-.29$); $0 = no$ relation ($r = -.09$ to $+.09$); $+1 = slight$ positive relation ($r = +.10$ to $+.29$); $+2 = a$ moderate or stronger positive relation ($r > +.30$)], which were averaged across raters and for each trait (see also O'Neill & Hastings, 2010).

Judges provided relevance ratings for six personality traits measured by the SFPQ (Jackson, Paunonen, & Tremblay, 2000): Agreeableness, Extraversion, Openness to Experience, Emotional Stability, Methodicalness, and Achievement. From that pool of six personality traits, we retained four traits because they exceeded an absolute relevance rating of 1.0. Concurrent validity would be supported if the Teamwork-KSA Test correlated, in the predicted direction, with the retained personality traits. We consider only the Teamwork-KSA overall score as it is the only score used in previous research, it was recommended for selection purposes (the other scores are recommended only for development and training; Stevens & Campion, 1993, 1999), and it was the only score sufficiently reliable to correlate with other variables. We report the observed relation, operational correlation corrected for unreliability in the personality variable, and the completely disattenuated validity. In this analysis, the power to detect a correlation of .20 was .86, and in order to reach significance the correlation needed to exceed .13.

Table 5. Concurrent validity of the Teamwork-KSA Test with theoretically related personality traits

Scale	Average judge relevance rating	Observed relation	Operational correlation ^a	Fully corrected validity ^b	Concurrent validity supported?
Agreeableness	1.67	.04	.05	.06	No
Achievement	1.17	.07	.08	.10	Yes
Emotional stability	1.17	-.01	-.01	-.01	No
Extraversion	1.33	.04	.05	.05	No

Notes: $N = 226$. ^aObserved correlation corrected for unreliability in the personality trait. ^bObserved correlation corrected for unreliability in the personality trait and the Teamwork-KSA overall scores. KSA = Knowledge, Skills, and Abilities.

Table 6. Discriminant validity of the Teamwork-KSA Test with psychological states and response distortion scales

Scale type	Variable	Observed relation	Operational correlation ^a	Fully corrected validity ^b	Discriminant validity supported?
Psychological state	Test-taking motivation ^c	.11	.13	.14	No
	Test-taking anxiety (W) ^d	-.22**	-.26	-.28	No
	Test-taking anxiety (E) ^d	-.13*	-.15	-.16	No
Response distortion	Self-deceptive enhancement ^e	.01	.01	.01	Yes
	Impression management ^f	.15*	.18	.20	No
Personality	Methodicalness ^g	-.03	-.04	-.05	Yes
	Openness to experience ^g	.13*	.14	.17	No

Notes: * $p < .05$, ** $p < .01$. $N = 266$. ^aObserved correlation corrected for unreliability in the psychological state, response distortion scale, or personality scale. ^bObserved correlation corrected for unreliability in the psychological state, response distortion scale, or the personality scale and the Teamwork-KSA Test. ^cTest-taking motivation is the extent to which respondents demonstrate effort, diligence, and focus in writing a test (Arvey et al., 1990). ^dTest-taking Anxiety can be bifurcated into two dimensions: Worry (TTA-W) and Emotionality (TTA-E) components (Liebert & Morris, 1967). TTA-W involves a concern about doing well on tests whereas TTA-E involves autonomic arousal, such as bodily tensions. ^eSelf-deceptive Enhancement is the extent to which an individual *unconsciously* responds in the favorably biased direction (Paulhus, 1991). ^fImpression Management is the extent to which an individual *intentionally* responds in the favorably biased direction. ^gThe average judge relevance rating for Methodicalness and Openness to Experience was .83 and .50, respectively. KSA = Knowledge, Skills, and Abilities.

As shown in Table 5, of the four traits that were rated as relevant to Teamwork-KSA Test scores by judges, only the fully corrected Achievement correlation was at a level that approached our judges' predictions. In addition, none of the correlations achieved significance. These findings provide limited concurrent validity evidence for the Teamwork-KSA Test as it relates to personality, which is surprising given that the KSAs falling under Stevens and Campion's (1994) taxonomy should, theoretically, be related to personality.

3.6. Discriminant validity

Discriminant validity of the Teamwork-KSA Test would be supported if the test was negligibly related to constructs that are theoretically unrelated to the Teamwork-KSA construct (see Murphy & Davidshofer, 2005; Walsh & Betz, 2001). To examine some evidence regarding the discriminant validity of the Teamwork-KSA Test, we considered its relations with psychological states related to taking tests (Test-taking Motivation and Anxiety) and Paulhus's (1991) response distortion scales (Self-deceptive Enhancement and Impression Management; for definitions and results, see Table 6). Ordinarily, tests' constructs are expected to be independent of the aforementioned variables – the hope is that tests measure the targeted attributes without

contamination of extraneous variance arising from Test-taking Motivation, Test-taking Anxiety, response distortion, or Impression Management (McCarthy & Goffin, 2003). We also investigated correlations with Methodicalness and Openness to Experience as these constructs were rated as unrelated to the Teamwork-KSA Test by our judges. Discriminant validity would be shown if negligible (i.e., $r = -.09$ to $+.09$) relations with the Teamwork-KSA Test were observed. We report the observed relations, operational correlations corrected for unreliability in the predictor variable (i.e., test-taking states, response distortion, personality), and the completely disattenuated discriminant validity coefficients. In this analysis, the power to detect a correlation of .20 was .91, and in order to reach significance, the correlation needed to exceed .12.

Table 6 indicates that, of the possible seven relations expected to be near zero in magnitude, four were significantly related to the Teamwork-KSA Test. In terms of effect sizes, five were small and two were negligible. Both Test-taking Anxiety-Worry and Test-taking Anxiety-Emotionality were negatively related to Teamwork-KSA Test scores. Moreover, Impression Management was positively related to the Teamwork-KSA Test. Thus, individuals' performance on the Teamwork-KSA Test is related, to some extent, with test anxiety and intentional response distortion (i.e., Impression Management). Finally, Open-

Table 7. Criterion correlations for the Teamwork-KSA Test

	1	2	3	4	5	6	7	8
1. Interpersonal KSAs								
2. Conflict resolution	.39							
3. Communication	.65	.23						
4. Collaboration and problem solving	.72	.18	.36					
5. Self-management KSAs	.49	.14	.44	.44				
6. Goal setting and performance management	.37	.10	.31	.33	.79			
7. Planning and task coordination	.41	.18	.41	.40	.77	.37		
8. Teamwork-KSA overall	.94	.32	.64	.69	.75	.58	.60	
9. Job performance	-.03	.02	.04	-.03	.03	.00	.04	.01
Operational ^a	(-.04)	(.03)	(.06)	(-.04)	(.04)	(.00)	(.06)	(.01)
Fully corrected ^b	(-.05)	(1.0) ^c	(.09)	(-.06)	(.06)	(.00)	(.09)	(.02)

Notes: Correlations above .13 are significant at $p < .05$. $N = 225$. Job performance is a composite of supervisor ratings of communication, collaboration, and teamwork. We estimated interrater reliability at .52 because this was the meta-analytic coefficient reported by Viswesvaran, Ones, and Schmidt (1996). ^aCriterion correlations corrected for unreliability in job performance. ^bCriterion correlations corrected for unreliability in job performance and the Teamwork-KSA Test. ^cThe near-zero reliability observed for the Conflict Resolution scale resulted in a spurious fully corrected correlation of 1.0 with job performance. KSA = Knowledge, Skills, and Abilities.

ness to Experience was positively related to the Teamwork-KSA Test, indicating some overlap with a trait rated as irrelevant for Teamwork-KSA Test performance by our judges. Taken as a whole, the discriminant validity analysis suggests that the Teamwork-KSA Test is relatively independent of variables to which it should be given that effect sizes were not more than small. Still, four small but significant correlations were detected where undesired (e.g., Test-taking Anxiety-Worry) or unexpected (e.g., Openness to Experience).

3.7. Criterion validity

One important component of the Teamwork-KSA Test's validity is its criterion validity. Given that core teamwork behaviors were identified in our job analysis, and that supervisors directly rated these dimensions, the Teamwork-KSA Test should be predictive in the present organization. Table 2 contains criterion validities for each item, although no items were significantly predictive of supervisor ratings of our team performance composite (communication, cooperation, and teamwork). Table 7 contains criterion validities for the Teamwork-KSA Test's overall, category, and scale scores as they relate to team performance ratings. The power to detect a correlation of .20 was .86, and in order to reach significance, the correlation needed to exceed .13. However, none of the observed correlations between the Teamwork-KSA Test scores and the composite team performance criterion approached statistical, or practical, significance. This does not support the conclusion that the Teamwork-KSA Test predicts team-related job performance competencies as seen by managers.

3.8. Post hoc analyses

In an attempt to salvage potentially psychometrically sound portions of the Teamwork-KSA Test, we deleted

items that could be contributing to poor reliability and, consequently, low concurrent and criterion correlations. Accordingly, we removed the seven items that had item-total correlations below .10 (see Table 2). As a result, Cronbach's alpha for the total Teamwork-KSA Test score increased from .71 to .75. Unfortunately, CFAs based on one, two, and five factors still did not result in acceptable model fits. EFAs also did not provide any interpretable structure, with the percentage of variation explained ranging from 10.97% for a one-factor solution to 27.83% for a seven-factor solution. The latter was based on a Parallel Analysis of the 28-item data set that suggested the retention of seven factors. Clearly, these psychometric properties do not meaningfully improve upon the ones reported earlier.

Concurrent correlations between the overall 28-item Teamwork-KSA Test score and personality did not sufficiently change to alter the conclusions reached in the existing analysis. That is, Emotional Stability, Agreeableness, and Extraversion all correlated below .10 with this adjusted Teamwork-KSA Test score, even after correcting all scales for unreliability. In addition, criterion correlations involving the adjusted Teamwork-KSA Test scores and job performance did not result in a different observed or fully corrected correlation (i.e., $\rho = .02$). Thus, deleting the items with the lowest item-total correlations from the Teamwork-KSA Test did not ameliorate its psychometric properties or validity coefficients in the present study.

4. Discussion

The movement toward evidence-based management discourages the application of applied practices that are incongruent with current research findings (Pfeffer & Sutton, 2006). The Teamwork-KSA Test is grounded in a strong theory-based taxonomy of teamwork capabilities

and is perhaps the most well-known, off-the-shelf test available for measuring such attributes. In order to make scientific advances in our understanding of the capabilities needed to effectively work in teams, by taking advantage of the solid and comprehensive theoretical framework provided by Stevens and Campion's (1994) taxonomy, a sound psychometric and construct valid measure is needed. The new empirical contributions of the present article, as well as the literature we reviewed, however, appear to be at odds with the citations in research and marketing suggesting that the Teamwork-KSA Test is a 'valid' instrument for selecting, developing, and training team members. This is where we see a significant gap in the current literature – some have claimed the Teamwork-KSA Test is valid (Hollenbeck et al., 2004), whereas our review of the literature and new analyses suggest the test's reliability and validity could be improved.

In our literature review, we uncovered four main areas where research is needed on the Teamwork-KSA Test: (a) reliability; (b) structural validity; (c) concurrent and discriminant validity; and (d) criterion validity. Accordingly, the goal of this study was to provide some evidence regarding each of these issues by summarizing the extant literature and presenting new empirical results. In what follows, we discuss our findings as they pertain to the four research needs identified earlier.

4.1. Reliability

To the extent that Teamwork-KSA Test scores are unreliable, decisions about applicants that are based on Teamwork-KSA Test scores will be less accurate. Decreased accuracy results in notably lower testing payoffs and possibly inappropriate decision making. For example, the standard error of measurement (SEM) can be computed to demonstrate the probability that two test scores are significantly different at less than a 5% level of chance (see Murphy & Davidshofer, 2005). The SEM for the Teamwork-KSA Test, based on our data is 2.48 [$SD \cdot \sqrt{1-r_{xx}}$]. Using standard test-banding techniques, this means that scores within 4.86 ($2.48 \cdot 1.96$) units above and below an individual's Teamwork-KSA Test score may not be reliably distinguished at the 95% level of confidence. For example, if one applicant receives a score of 33 (the test score maximum in the current data), other applicants may not be viewed as performing reliably poorer unless their test scores were equal to or lower than 28. This illustrates how the utility of the test is impacted by reliability. Note that category- and scale-level scores would have much larger confidence bands, thereby limiting their discriminating capabilities. This is an issue given the recommended use of category and scale scores for development and training (see Stevens & Campion, 1993, 1999), which are consequential for both individuals and organizations.

How do the current reliability coefficients compare with previous research? Stevens and Campion (1999) deleted two unidentified items with negative item-total correlations to arrive at their overall test reliability of .80, suggesting the full-test reliability was potentially less than .80. Elsewhere, studies have reported lower test-score reliabilities for the overall test score (e.g., $\alpha = .59$; McClough & Rogelberg, 2003). Leach et al. (2005) reported that they could not investigate category and scale relationships because of exceedingly low reliabilities. Taken together, reliability of the overall Teamwork-KSA Test score appears to be in the acceptable range, although perhaps a little lower than some other selection tests such as integrity and GMA reviewed earlier, but scale and category scores may be too unreliable for most applications. We interpret the current reliability analysis as suggesting that there is room for improvement in the Teamwork-KSA Test's reliability, particularly at the lower level, if it is to be applied to many areas of personnel management.

4.2. Structural validity

Our assessment of structural validity through a variety of standard and advanced factor-analytic methods further underlines reliability concerns. Indeed, no interpretable factor structure could be recovered. We urge future researchers to conduct and report factor-analytic findings regarding the Teamwork-KSA Test. Evidence of structural validity is imperative to support the inferences made based on scores at the various levels of the Teamwork-KSA Test hierarchy.

4.3. Concurrent and discriminant validity

Concurrent and discriminant validity evidence, as evaluated in our study, was mixed. Concurrent validity coefficients with personality traits judged to be relevant to Teamwork-KSA Test scores were all in the very small-to-trivial range even when completely disattenuated, and none were statistically significant. Discriminant validity was generally supported in that the Teamwork-KSA Test did not correlate strongly with test-taking states, response distortion scales, or Methodicalness. However, the relation with Test-taking Anxiety – Worry was low to medium in magnitude ($\rho = -.28$). This suggests that worrying about performing well on the test was detrimental to Teamwork-KSA Test performance. Normally, it is undesirable to have anxiety impact test scores, and future research could consider whether this finding generalizes to other settings and test-taking situations.

Recall from the literature reviewed earlier that the Teamwork-KSA Test generally correlated higher with taskwork than it did with teamwork criteria. An

explanation for the stronger trend toward taskwork predictiveness by the Teamwork-KSA Test is that the Teamwork-KSA Test is a surrogate for GMA, which is expected to be more important for taskwork than it is for teamwork. Supportively, the correlation of the Teamwork-KSA Test with GMA was found to be .81, which, corrected for measurement error, becomes .91 and .99 in the two samples reported by Stevens and Campion (1999). Accordingly, practically significant incremental validity beyond standard measures of GMA would be unlikely. More generally, these findings beg the question as to whether the Teamwork-KSA Test is truly a construct valid measure of teamwork KSAs rather than of GMA.

4.4. Criterion validity

The criterion validity of the Teamwork-KSA Test was not supported even though the present study occurred in a team-based organization with a well-developed criterion highly relevant to teamwork. This adds to the mixed results reported earlier and by previous authors (Allen & West, 2005), but also makes a significant and unique contribution by considering criterion relations among all categories and scales. In other analyses not reported here, we did investigate the incremental validity of the Teamwork-KSA Test beyond the six SFPQ personality variables, but incremental prediction was not supported. Overall, given the importance of criterion validity in the legal defensibility of the test, we see use of the Teamwork-KSA Test as a concern for its ongoing application in applied practice without more validation work and further development.

4.5. Limitations and strengths

It was not possible to meta-analyze the existing criterion correlations on the Teamwork-KSA Test. Studies are scattered across contexts, types of criteria, and are mixed between team and individual levels of analysis. Note that the issue of too few studies does not mean the Teamwork-KSA Test is unpopular; on the contrary, we know of no other standardized teamwork selection device, developed in the research literature, that was subsequently referenced by numerous scholars and continues to be marketed to organizations as a valid selection tool. Accordingly, we conducted a detailed review of existing literature (see Table 1), and our review exposed the need for additional evidence on the psychometric properties and validity of the Teamwork-KSA Test. Our empirical study, conducted in an organizational context highly appropriate for studying teamwork variables, sought to address many of the novel concerns uncovered in the literature review.

A limitation to the results of our empirical study is that it is a single-sample investigation. This brings with it the

usual generalizability concerns. However, our sample was not composed of a single employment position; we included a very wide range of team-oriented jobs in order to increase generalizability across occupations. Moreover, we rigorously developed our teamwork criterion using multimethod job analysis conducted in a team-based organization to ensure that the competencies advanced in Stevens and Campion's (1994) taxonomy were well aligned to those required for the present jobs. Our sample comprised some jobs that were similar to those used in previous field studies on the Teamwork-KSA Test, even those supportive of the test's validity (e.g., Stevens & Campion, 1999; Morgeson et al., 2005). Statistical power was not a concern as it was substantial, and all correlations exceeding .13 could be detected as statistically significant. Finally, our post hoc analyses not including items with low item-total correlations did not improve the outlook for the Teamwork-KSA Test.

6. Conclusion

In this article, we present new evidence regarding the reliability and validity of the Teamwork-KSA Test. Some previous studies have either not reported reliability coefficients (e.g., Miller, 2001), or have reported low reliabilities (e.g., $\alpha = .59$; McClough & Rogelberg, 2003). McClough and Rogelberg suggested the low reliability was due to the multidimensionality of the overall test; however, our analysis reveals that the narrower scales appear to fare far worse, such that no reliabilities exceeded even .50 at the scale level and one had a reliability of zero. Also, reflective of weak covariance among like items, our CFA and EFA, the first of our knowledge to be presented in the published literature, raise serious questions regarding the extent to which the overall Teamwork-KSA Test score, and especially the scale scores, can be meaningfully interpreted.

The Teamwork-KSA Test has been advertised as a test that can be used for selection-, promotion-, and training-related needs (Creative Organizational Design, 2009), yet the new evidence reported here suggests that important decisions based on Teamwork-KSA Test scores might be suboptimal. Of course, our study is the only one, to our knowledge, that has investigated the factor structure of the Teamwork-KSA Test and that has reported item- and scale-level psychometric properties. More research of this type is sorely needed in order to arrive at an overall consensus regarding the viability of the Teamwork-KSA Test.

Note

1. We treated Stevens and Campion's two samples separately in the aggregation; therefore, we did not include their combined-sample findings in the current analysis.

References

- Aguinis, H., Henle, C. A., & Ostroff, C. (2001). Measurement in work and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology 1*, (pp. 27–50). London: Sage.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth Inc.
- Allen, N. J., & West, M. A. (2005). Selection for teams. In A. Evers, N. Anderson, & O. Voskuil (Eds.), *Handbook of personnel selection* (pp. 476–494). Oxford: Blackwell Publishing.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695–716.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Bentler, P. M. (2003). *EQS 6.1 for Windows*. [Computer Software]. Encino, CA: Multivariate Software.
- Bobko, P. (2001). *Correlation and regression* (2nd ed.) Thousand Oaks, CA: Sage.
- Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning and Education*, 3, 27–40.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 115–159.
- Creative Organizational Design. (2009). Teamwork – knowledge, skills, attitudes. Available at <http://creativeorgdesign.com/ksa.htm> (accessed December 4, 2009).
- Cronbach, L. J., & Meehl, P. E. (1978). Construct validity in psychological tests. In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 57–77). Huntington, NY: Robert E. Krieger Publishing Company.
- Ellis, A. P. J., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, 58, 641–672.
- Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles, CA: Sage.
- Goffin, R. D. (1993). A comparison of two new indices for the assessment of fit of structural equation models. *Multivariate Behavioral Research*, 28, 205–214.
- Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology*, 11, 23–33.
- Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The Relative Percentile Method. *Human Resource Management*, 48, 261–268.
- Graen, G. B. (2009). Educating new management specialists from an evidence-based perspective: A proposal. *Academy of Management Learning and Education*, 8, 255–258.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Harvey, R. J. (1993). *Research monograph: The development of the CMQ. Monograph describing the development and field-testing of the Common Metric Questionnaire (CMQ)*.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7, 191–205.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21, 967–988.
- Hirschfeld, R. R., Jordan, M. H., Feild, H. S., Giles, W. F., & Armenakis, A. A. (2006). Becoming team players: Team members' mastery of teamwork knowledge as a predictor of team task proficiency and observed teamwork effectiveness. *Journal of Applied Psychology*, 91, 467–474.
- Hollenbeck, J. R., DeRue, D. S., & Guzzo, R. (2004). Bridging the gap between I/O research and HR practice: Improving team composition, team training, and team task design. *Human Resource Management*, 43, 353–366.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus narrow alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jackson, D. N. (1999). *Personality research form, form I*. Port Huron, MI: Sigma Assessment Systems.
- Jackson, D. N., Ashton, M. C., & Tomes, J. L. (1996). The six-factor model of personality: Facets from the big five. *Personality and Individual Differences*, 21, 391–402.
- Jackson, D. N., Paunonen, S. V., Fraboni, M., & Goffin, R. D. (1996). A five-factor versus six factor model of personality structure. *Personality and Individual Differences*, 20, 33–45.
- Jackson, D. N., Paunonen, S. V., & Tremblay, P. F. (2000). *Six factor personality questionnaire manual*. Port Huron, MI: Sigma Assessment Systems.
- Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95, 54–78.
- Kline, P. (1984). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen, Inc.
- Kline, R. B. (2004). *Principles and practices of structural equation modeling* (2nd ed.) New York: Guilford.
- Kottke, J. L. (2008, April). *Able but not willing? Teamwork aptitude and interest meet head-on*. Paper presented at the 23rd Annual Meeting of the Society for Industrial and Organizational Psychology, San Francisco.
- Leach, D. J., Wall, T. D., Rogelberg, S. G., & Jackson, P. R. (2005). Team autonomy, performance, and member job strain: Uncovering the teamwork KSA link. *Applied Psychology: An International Review*, 54, 1–24.
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61, 273–307.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20, 975–978.
- Loevinger, J. (1967). Objective tests as instruments of psychological theory. In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 78–123). New York: McGraw-Hill. Reprinted from *Psychological Reports*, 1957, Monograph Supplement).
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26, 356–376.
- Martín-Pérez, V., Martín-Cruz, N., Pérez-Santana, P., Hernangómez-Barahona, J., & Martín-Sierra, C. (2010).

- 'Virtual learning-by-doing' teamwork KSA: Strategic management simulation as an effective tool. *Communications in Computer and Information Science*, 73, 324–330.
- McCarthy, J. M., & Goffin, R. D. (2003). Is the Test Attitude Survey psychometrically sound? *Educational and Psychological Measurement*, 63, 446–464.
- McCarthy, J. M., & Goffin, R. D. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology*, 57, 607–637.
- McCarthy, J. M., & Goffin, R. D. (2005). Selection test anxiety: Exploring tension and fear of failure across the sexes in simulated selection scenarios. *International Journal of Selection and Assessment*, 13, 282–295.
- McClough, A. C., & Rogelberg, S. G. (2003). Selection in teams: An exploration of the teamwork knowledge, skills, and ability test. *International Journal of Selection and Assessment*, 11, 56–66.
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, 107, 242–255.
- McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology*, 36, 979–1016.
- Messick, S. (1991). *Validity of test interpretation and use*. Princeton, NJ: Research Report for the Educational Testing Service.
- Miller, D. L. (2001). Reexamining the teamwork ksas and team performance. *Small Group Research*, 32, 745–766.
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, 58, 583–611.
- Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, 11, 145–165.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. C., & Burnstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396–402.
- O'Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2010). Test-taking motivation and personality test validity. *Journal of Personnel Psychology*, 9, 117–125.
- O'Neill, T. A., & Hastings, S. E. (2010). Explaining workplace deviance behavior with more than just the 'Big Five'. *Personality and Individual Differences*, 50, 268–273.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality*, 17, S19–S38.
- Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. (1983). Statistical power of tests of situational specificity hypothesis in validity generalization studies: A cautionary note. *Journal of Applied Psychology*, 68, 115–122.
- Oswald, F. L., & Johnson, J. W. (1998). On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: Some initial Monte Carlo findings. *Journal of Applied Psychology*, 83, 164–178.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychology attitudes* (pp. 17–59). New York: Academic Press.
- Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, 84, 62–74.
- Ramsay Corp. (2009). Teamwork KSA. Available at <http://www.ramsaycorp.com/products/teamworkksa.asp> (accessed December 4, 2009)
- Rapp, T. L., & Mathieu, J. E. (2007). Evaluating an individually self-administered generic teamwork skills training program across time and levels. *Small Group Research*, 38, 532–555.
- Rousseau, D. (2006). Is there such a thing as evidence-based management? *Academy of Management Review*, 31, 256–269.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573–605.
- Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629–637.
- Schulte, M. J., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: Not much more than g and personality. *Personality and Individual Differences*, 37, 1059–1068.
- Spielberger, C. D. (1980). *Test attitude inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C. D., & Vagg, P. R. (1995). *Test anxiety: Theory, assessment, and treatment*. Washington, DC: Taylor and Francis.
- Stevens, M. J., & Campion, M. A. (1993). *The Teamwork KSA test*. Minneapolis, MN: NCS Pearson, Inc.
- 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.
- Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, 25, 207–228.
- Stevens, M. J., Jones, R. G., & Fisher, D. L. (2002, April). *What's past is prologue: Exploring a biodata approach to team selection*. In F. P. Morgeson (chair), Selection for Teams: A tale of five approaches. Symposium presented at the 17th Annual Meeting of the Society for Organizational and Industrial Psychology, Toronto, Ontario, Canada.
- Stewart, G. L., Fulmer, I. S., & Barrick, M. R. (2005). An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. *Personnel Psychology*, 58, 343–365.
- Taggar, S., & Brown, T. C. (2001). Problem-solving team behaviors: Development and validation of BOS and a hierarchical factor structure. *Small Groups Research*, 32, 698–726.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at Seventy* (pp. 42–72). Boston, MA: Kluwer Academic Publishers.
- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in 'Big Five Factors' personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60, 224–235.

- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 5, 557–574.
- Wagner, S. H., & Goffin, R. D. (1997). Differences in accuracy of absolute and comparative performance appraisal methods. *Organizational Behavior and Human Decision Processes*, 70, 95–103.
- Walsh, W. B., & Betz, N. E. (2001). *Tests and assessment* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- West, M. A., & Allen, N. A. (1997). Selecting for teamwork. In N. Anderson & P. Herriot (Eds.), *International Handbook of Selection and Assessment* (pp. 493–506). New York: John Wiley.
- Williams, H. M., & Allen, N. J. (2008). Teams at work. In J. Barling & G. L. Cooper (Eds.), *The Sage handbook of organizational behavior*. (Vol.1, pp. 124–140). Los Angeles, CA: Sage.