Journal of Education for Business

ISSN: 0883-2323 (Print) 1940-3356 (Online) Journal homepage: http://www.tandfonline.com/loi/vjeb20

Teamwork skill assessment: Development of a measure for academia

Otmar Varela & Esther Mead

JOURNAL OF EDUCATION FOR BUSINESS
https://doi.org/10.1080/08832323.2018.1433124

INNOVATIVE INSTRUCTIONAL CLASSROOM PROJECTS/BEST PRACTICES

Teamwork skill assessment: Development of a measure for academia

Otmar Varela[a] and Esther Mead[b]

[a]Department of Management, University of Arkansas-Little Rock, Little Rock, Arkansas, USA; [b]Information Sciences, University of Arkansas-Little Rock, Arkansas, USA

ABSTRACT
Popular teamwork assessments have been strongly criticized on the grounds of poor psychometric properties and their disconnect with conceptual models of teamwork. These issues raise concerns with respect to our ability to evaluate efforts devoted to advancing teamwork in academia. We report the development of a teamwork assessment that builds on empirically supported conceptualizations of team processes. Two studies were conducted to test and to cross-validate the psychometrics of the resulting measure. In the discussion section, we address the implications of our findings for conceptual models of teamwork and provide guidelines for using the measure in business education.

KEYWORDS
Classroom assessment; scale development; teamwork

It is difficult to overstate the importance of teams for organizations. In 2015, 95% of U.S. employees reportedly worked in more than one team (Center for Creative Leadership, 2015). Similarly, the European Observatory of Working Life reports a high incidence of teams in organizations, with the United Kingdom exhibiting the highest incidence in a sample of 16 countries (80.6% of jobs; European Observatory of Working Life, 2007). Thus, it is unsurprising to find consensus in defining teamwork as a core competence in models that define the employability of higher education graduates (e.g., Betta, 2016).

Something more contentious is the effectiveness of pedagogies directed at developing teamwork skills, with meta-analyses showing a dispersion of results (McEwan, Ruissen, Eys, Zumbo, & Beauchamp, 2017; Salas et al., 2008). Many reasons may explain the fluctuation of outcomes (e.g., personal attributes, unfitted pedagogies). Our key concern deals with the validity of assessments used in quantifying results. Most of studies that address narrowly defined teamwork dimensions (e.g., communication) exhibit assessments with sound psychometric properties (e.g., Valls, Gonzalez-Roma, & Tomas, 2016). However, studies addressing a more comprehensive domain of the skill seem to face significant challenges. In academia, for example, rubrics (the modal assessment) list several teamwork dimensions detailed in behaviorally anchored rating scales that facilitate instruction and feedback (Jonsson & Svingby, 2007). And although an abundance of rubrics exists, we were unable to locate a study that reports rubrics with strong psychometric properties. Rather, studies are mute with respect to the internal factor structure or reliabilities of these assessments (e.g., Cela-Ranilla, Esteve-Mon, Esteve-Gonzalez, & Gisbert-Cervera, 2014; Kemery & Stickney, 2014), a circumstance that raises questions with respect to the accuracy of learning outcomes achieved in business education and other fields (e.g., medical education; see Lerner, Magrane, & Friedman, 2009).

The assessments used in training are not strange to similar criticisms. Stevens and Campion's (1999) Teamwork KSA, a measure of extensive use in the industry, has been repeatedly criticized for its poor psychometric properties (O'Neill, Goffin, & Gellatly, 2012). In their study, Aguado, Rico, Sanchez-Manzanares, and Salas (2014) concluded that Stevens and Campion's assessment "does not adequately reflect the initial substantive model [of teamwork] and has limitations with regard to reliability" (p. 101). The psychometrics of other popular assessments have been similarly the target of significant concerns (e.g., Team Development Survey: Hallam & Campbell, 1997; Teamwork Quality: Hoegl & Gemuenden, 2001), with warnings indicating these assessments lead to suboptimal decisions (Senior & Swailes, 2007).

Another controversy with respect to existing measures of teamwork deals with their content validity with assessments exhibiting discrepancies in their underlying factor structures. While several dimensions (e.g., communication,

CONTACT Otmar Varela oxvarela@ualr.edu Management Department, University of Arkansas-Little Rock, 2801 South University Ave, Little Rock, AK 72204, USA.

coordination) are recurrently cited as core components (e.g., Hoegl & Gemuenden, 2001), disparities emerge beyond such central dimensions. In a review of the literature, Rousseau, Aube, and Savoie (2006) listed 29 variations in frameworks of teamwork. These variations include the number of dimensions, degree of specificity, and hierarchical arrangements, all symptoms of a lack of consensus that prevails in the literature. We echo Aguado et al. (2014) in questioning the extent to which existing measures exhaustively tap the behavioral dimensions necessary to work in teams.

There is an urgent need to advance assessments that validate the enormous educational efforts devoted to teamwork. This development must rely on current theorizing and empirical evidence on the factor structure of teamwork. This article describes two studies directed at meeting this goal. To this end, we first review the domain of teamwork. The ensuing section describes the methods utilized in assembling the measure and the studies conducted for testing its psychometric properties. The final section reflects on the theoretical implications of the assessment and provides guidelines for using the measure in academia.

Teamwork

Team denotes a group of two or more people working interdependently in the pursuit of a common goal. This prevalent notion of team exhibits three features shown at the core of most conceptualizations: (a) commonality of goals across members, (b) synergy that emerges from members' interdependence, and (c) size, with at least two members viewed as sufficient (Hare, 2010). Paradoxically, something that has received less attention and consequently is more contentious is the notion of teamwork. An overview of the literature indicates discrepancies with respect to the nature, boundaries, and factor structure of the construct (Salas, Sims, & Burke, 2005). For example, while teamwork can be viewed as a set of attributes necessary to functioning in teams (e.g., knowledge, flexibility), it can also address interpersonal actions that facilitate collective work (e.g., communication). That is, teamwork is simultaneously used to capture both personal traits and a set of interpersonal behaviors necessary for teams to operate. For purpose of this study, we adopt the behavioral view and define teamwork as a set of collective actions instrumental to generate valuable results for teams (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008).

A fundamental reason for adopting the behavioral view of teamwork is the utility of the assessment: validating learning. Our goal is to generate a measure that judges a resulting capability to work in teams. It is our position that reflective indicators, via behaviors, better capture the degree of skill proficiency while facilitating assessments from distinct sources (e.g., peers), something more contentious in measures that target attributes (e.g., knowledge). Attribute-based assessments not only are inferential, as they assume behavioral capabilities from a set of underlying attributes, but also may be deficient, as we are unaware of an exhaustive list of attributes that may account for the universe of teamwork behaviors. We are not arguing for the irrelevance of attribute-based assessments. They are of significant value in situations that involve forecasting (e.g., selection). However, because our goal is to validate outcomes in learning settings, we seek for indicators that clearly reflect a learned capability to work in teams.

We draw from the team processes literature to identify the domain of teamwork. McGrath's (1984) input-process-output model is arguably the prevalent view of team effectiveness, and as such experts have emphasized team processes in establishing the actions that lead to team success (Mathieu, Gilson, & Ruddy, 2006). Specifically, processes are collective contributions that transform inputs (e.g., knowledge) into valuable results. Analysis of processes is instrumental to identify meaningful members' actions. As Baker and Salas (1992) stated, group processes "provide the basis for developing the theoretical underpinnings of teamwork" (p. 471).

We adopt Marks, Mathieu, and Zaccaro's (2001) processes framework as a starting model. The framework emerged from an extensive review of the literature and, to the best of our knowledge, is the only one exhibiting meta-analytical support (LePine et al., 2008). The framework lists 10 lower-level dimensions hierarchically arranged into three superordinate categories: Transition processes, action processes, and interpersonal processes. A unique feature of the model is the consideration of time. Processes are viewed episodically, in the sense that they unfold at different stages contingent on the phase of task completion.

Transition processes occur at early stages of task development. They set guidelines for future action by establishing the boundaries of the tasks and defining avenues for goal achievement. Marks et al. (2001) associated three behaviors with transition processes—mission analysis, goal specification, and strategy formulation. Collectively, these actions define the purpose of the team, grain the mission into subgoals, and identify methods for achieving the team's purpose. Table 1 includes a brief definition of each category. Readers are referred to Marks et al. for an extensive definition.

Action processes advance the tasks. These actions are interpersonal by nature and should be separated from technical-job tasks (e.g., manipulating tools). As Marks et al. (2001) noted, "Taskwork represents *what* it is that teams are doing, whereas teamwork [action processes]

Table 1. Marks et al.'s (2001) classification of team processes.

Broad Category	Lower-level dimension	Definition
Transition processes	Mission analysis	Definition of the team main task and appraisal of resources to accomplishing mission.
	Goal specification	Articulation of team goals derived from mission analysis and prioritization of more specific goals.
	Strategy formulation	Establishing courses of action for mission accomplishment.
Action processes	Monitoring progress	Ensuring the task advances according to pre-established goals and procedures.
	System monitoring	Ensuring that existing resources and conditions suffice for goal accomplishment.
	Team monitoring and backup behaviors	Assisting other members in performing their tasks.
Coordination		Orchestrating teammates' actions toward goal accomplishment
Interpersonal processes	Conflict management	Preventing harmful conflict or finding healthy resolution for existing conflicts
	Motivating and confidence building	Supporting the spirit of the team. Reinforcing team cohesiveness.
	Affect management	Regulating emotions that may interrupt advancement of collective tasks.

describes *how* they are doing it with each other" (p. 357, italics added). Four behaviors are listed under action processes: monitoring progress toward goals, system monitoring, backup responses, and coordination. Cannon-Bowers and Bowers (2011) suggested that two of these behaviors, monitoring task progress and system monitoring, can be subsumed into a single category, as their common tracking nature makes them hard to separate in practice. While we follow the original Marks et al. proposition as blueprint for the measure, we return to the Cannon-Bowers and Bowers suggestion later. Table 1 includes a brief definition for each of the action processes.

Interpersonal processes refer to actions that govern the quality of teammates' relationships. They fundamentally generate the so-called *esprit de corps*—collective attitudes that drive loyalty and feelings of pride towards the team. Interpersonal processes are the primary antecedents in studies that explore the social fabric of groups (e.g., Mortensen & Hinds, 2001). Marks et al. (2001) listed three behaviors as interpersonal processes: conflict management, motivating others and confidence building, and affect management, and Table 1 shows a summary of their content.

Mathieu et al. (2006) generated a survey that targets the three higher-level dimensions listed in Table 1. Studies in academic settings have similarly attempted to assess this triadic view of team processes (e.g., Bravo, Lucia-Palacios, & Martin, 2016). Yet, to the best of our knowledge, existing surveys ignore the lower-level dimensions, thus limiting specificity in the behavioral categories. Our goal is to develop an assessment that targets the entire structure in Table 1.

Methods

We followed Hinkin's (1995) suggestions for best practices in scale development. Accordingly, we (a) generated a pool of items that mirrors the structure in Table 1, (b) tested the items' ability to convey meaningful information, (c) examined the internal factor structure of the scale, (d) assessed the scale's convergent and discriminant validity, and (e) corroborated the stability of the resulting instrument across raters. In the following sections we detail each step.

Generating a pool of items

To ensure content validity (Hinkin, 1995), our first task consisted of generating items that reflect the structure in Table 1. Rather than creating items deductively, we conducted a literature review in search of existing measures addressing similar dimensions. Our rationale was that items from existing measures possess a proven relationship with their underlying dimension. As such, two research assistants with teamwork experience along with Otmar Varela (subject matter experts) sorted items from the following measures: the teamwork expectations scale (Eby, Meade, Parisi, & Douthitt, 1999), the teamwork survey (Senior & Swailes, 2007), the teamwork evaluation form (Hobson & Kesic, 2002), teamwork in higher education (Viles, Zarraga-Rodriguez, & Jaca, 2013), and team processes (Mathieu et al., 2006).

To select items from measures, we observed their factor loadings (when reported) and carefully considered if their wording unequivocally evokes a teamwork dimension in Table 1. Items were only selected by consensus. This process led us to adopt 21 items, most of which were adapted to fit a statement form and to target individuals rather than the entire team. For example, we modified the item "team roles are clearly defined" from Senior and Swailes's (2007) measure into "this person understands his/her team role." Because our goal was to generate a scale with at least three items per dimension, subject matter experts developed items deductively for those dimensions that remained with two items or less after sorting the items from the previous measures.

The resulting scale consisted of 35 items, with five of the 10 teamwork dimensions exhibiting four items. The scale was developed in a transparent format where a definition of the dimension antecedes the corresponding set of items. Instructions ask respondents to assess the accuracy of the behavioral statements for which a 7-point Likert-type scale with responses ranging from 1 (*strongly disagree*) to 7 (*strongly agree*) was added. We submitted the measure to a panel who was asked to provide feedback with respect to clarity of interpretation. This panel consisted of 10 graduate students attending a management course. Because the scale intends to assess teamwork in academic settings, this panel was considered appropriate. Minor modifications followed feedback from the panel.

Study 1: Testing items' performance

Participants and Procedures

Participants of the study consisted of 176 business graduate students of a southern U.S. University. Average age of the sample was 29.6 years old ($SD = 4.98$ years), with 41% being women. Participants were enrolled in various graduate courses that require teamwork. They were asked to complete a self-assessment of teamwork with respect to their experiences across the different teams. The survey was applied towards the end of the 14-week semester. Instructions of participation, detailed in the institutional review board consent form, made abundantly clear that participation was voluntary, data were collected anonymously, and results of assessments would not interfere in any way with course grading. Only two participants of the original 178 who were invited to participate declined the invitation.

Performance of Items

Our first test consisted of exploring the items' ability to discriminate between raters by observing items' skewness and standard deviation. As Clark and Watson (1995) stated, items with highly skewed distribution and low standard deviation "convey little information … fare poorly in subsequent structural analysis and … produce highly unstable correlation results" (p. 315). Of note, we expected asymmetries in the items' distributions. The specialized literature abundantly speaks to the significant biases in self-assessments (e.g., Dunning, Heath, & Suls, 2004), particularly when the target is a competence of value (teamwork).

Results corroborated our expectation in the sense that all items exhibited negative skewness (skewness average $= -0.86$, $SD = 0.21$). A rule of thumb is that skewness above ± 1 indicates extreme asymmetries (e.g., Doane & Seward, 2011), and as such items surpassing these statistics must be carefully re-examined. In addition to the ± 1 skewness threshold, we also observed the items' standard deviation as an indication of their ability to discriminate among raters. The average standard deviation of items was 0.94 ($SD = 0.20$). Given these results, a decision was made to drop items exhibiting skewness of <-1, and SD < 0.74 in which the latter represents one standard deviation below the items' average. These criteria led us to drop two items from the original list of 35.

Factor structure of the measure

The remaining set of 33 items was submitted to confirmatory factor analysis (CFA) using (Multivariate Software, Inc. manufactures EQS 6.1, Los Angeles, CA). The underlying factor structure was specified as outlined in Table 1. Maximum likelihood estimation was used as a method for parameter calculations. Three CFA outputs were of particular interest: general fit statistics, standardized residuals, and Lagrange multiplier test (LMT). While general fit statistics indicate the ability of the underlying model (Table 1) to account for the study data, the other two outputs inform on the items' performance in the context of the hypothesized structure. The model was arranged hierarchically with three second-order factors (transition, action, and interpersonal processes) in the apex. Consistent with theorizing on the interdependence of teamwork processes (LePine et al., 2008), second-order factors were allowed to correlate.

Table 2 exhibits results of the CFA. Two statistics—comparative fit index (CFI) and incremental fit index (IFI)—contrast the hypothesized model to a null one in which variables are assumed to be uncorrelated. CFI and IFI were selected for their parsimony with respect to sample size and degrees of freedom (Byrne, 2006). Values of > 0.90 typically indicate appropriate model fit. We also observed the root mean square error of approximation (RMSEA), a statistic that tests the hypothesized model against the study data. RMSEA reports an aggregate of standardized residuals in a metric that is sensitive to the number of estimated parameters, thus addressing parsimony with respect to degrees of freedom. Values of < 0.10 typically indicate appropriate model fit. Following conventions, we also include chi-square results in Table 2.

Table 2. Confirmatory factor analyses of teamwork measure.

Model	Sample	χ^2	*df*	CFI	IFI	RMSEA	$\Delta\chi^2$
1	Self-assessment	970.60	479	0.824	0.827	0.077	
2	Self-assessment	452.50	259	0.906	0.908	0.065	518.10*
3	Self-assessment	382.60	236	0.924	0.925	0.060	69.90*
4	Peer assessment	448.86	236	0.914	0.916	0.072	
5	Peer assessment	432.79	235	0.924	0.925	0.068	16.10*
6	Multigroup	818.83	471	0.924	0.926	0.064	
7	Multigroup	836.84	448	0.916	0.917	0.066	18.01

Note. CFI = comparative fit index; IFI = incremental fit index; RMSEA = root mean square error of approximation.
*$p < .05$.

However, given the sensitivity of chi square to sample size (Jöreskog & Long, 1993), this statistic is exclusively utilized to compare alternative models.

Results in Table 2 (Model 1) indicate that the hypothesized model replicates the study data inadequately. Although RMSEA (0.077) met accepted standards, CFI (0.824) and IFI (0.827) fell short of conventions, broadly suggesting the measure needed adjustments. To do so, we first explored CFA outputs assessing individual parameters. Our goal was to examine the performance of items via standardized residuals. The assumption here is that large residuals are indicative of items' misfit, particularly with respect to the hypothesized model. Because residuals are standardized, values of > 1.96 were considered alerting (Jöreskog & Long, 1993). Results indicated that none of the individual parameters surpassed the 1.96 threshold. Standardized residuals ranged from -0.191 to 0.605 (average $= 0.064$). We interpreted this outcome as evidence that the items perform appropriately in the context of the hypothesized model, and as such we sought for other sources of model misfit better explaining the previous statistics.

The LMT informs of model modifications that improve model fit. Our key concern, however, was not to capitalize on these modifications by chance. To avoid opportunistic adjustments, we sought for modifications consistent with existing theorizing on teamwork dimensions. As mentioned previously, one of the alternatives to our original model is Cannon-Bowers and Bowers's (2011) review of teamwork. While they coincided with Marks et al. (2001) in viewing teamwork as a triad hierarchical structure, Cannon-Bowers and Bowers suggested a simplification of the lower-level dimensions. Specifically, Cannon-Bowers and Bowers placed system monitoring and monitoring progress into a single dimension and suggested a similar merging between goal specification and strategy formulation. In the case of system monitoring and monitoring progress, the merging is a consequence of perceptions of overlap. Similarities in the nature of both behaviors—tracking—may lead raters to seen them as equivalent. Although the target of actions differs (resources in system monitoring and the task in monitoring progress), one could argue difficulties to separate them in applied settings and, as such, "these behaviors are rarely studied in isolation" (Cannon-Bowers and Bowers, 2001, p. 619). Similarly, the behaviors under mission analysis and goal specification can be perceived as duplication, as fundamentally both dimensions seek to define the boundaries of team objectives.

Moreover, LePine et al.'s (2008) meta-analysis, arguably the key empirical evidence on the hierarchical structure in Table 1, primarily concentrates on exploring the higher-level dimensions' ability to account for relevant outputs (i.e., performance and members' satisfaction). Yet, a direct test on the exhaustiveness or conversely the overlap between lower-level dimensions was overlooked. This is surprising, especially given the between-factor correlations LePine et al. report. In their study, the correlation between system monitoring and team monitoring along with the correlation between goal specification and mission analysis are noticeably high (.77 and .74, respectively). On the grounds of psychometric theory (Nunnally, 1978), one could argue that correlations of this magnitude suggest that factors are tantamount expressions of the same construct, a clear indication of the overlap among lower-level dimensions. Additionally, in our results, the LMT statistics associated with Model 1 further indicate the overlap between the same dimensions. Specifically, LMT shows the addition of a link between mission analysis and goal specification, and a link between system monitoring and monitoring progress are two of the top three suggested modifications.

We proceeded to testing the eight-factor solution with three higher-order dimensions (Model 2). To select items for the merged factors, we retained the three items exhibiting lowest residuals; that is, the ones indicating superior model fit. Results indicate that Model 2 provides a stronger model fit when compared with Model 1 (See Table 2). Not only do the fit statistics meet accepted standards (CFI $= 0.906$, IFI $= 0.908$, RMESA $= 0.065$), but also a direct comparison to Model 1 indicates statistically significant improvements, $\Delta\chi^2(220) = 518.10$, $p < .05$. Of note, Model 2 was tested with a 27-item measure in which only three factors (affect management, motivating others, and strategy formulation) included four items.

For parallel purposes and to minimize raters' fatigue (Hinkin, 1995), we dropped one item from each of the 4-item factors so that the resulting measure included three items across dimensions. We observed the items' standardized residuals in identifying items to be dropped. Following LMT suggestions from Model 2, we also allowed for a double loading on one of the items (item 22 in Appendix A). Model 3 in Table 2 exhibits results of the CFA when the 24-item measure was utilized. Outcomes (Model 3) show that the 24-item measure provides a better fit with adequate fit statistics (CFI $= 0.924$, IFI $= 0.925$; RMSEA $= 0.060$) and a $\Delta\chi^2$ indicating the superiority of this option, $\Delta\chi^2(23) = 69.90$, $p < .05$. Appendix A exhibits the resulting 24-item instrument. Table 3 exhibits the between factor correlations along with Cronbach's alpha reliabilities for each of the eight dimensions in the diagonal.

Confirmatory and discriminatory tests

Confirmatory and discriminatory analyses were conducted by observing the pattern of correlations among

Table 3. Correlations among teamwork dimensions, personality traits, and cognitive abilities.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. Mission analysis	6.15	0.56	(.81)									
2. Strategy formulation	5.76	0.75	.63	(.90)								
3. Situation monitoring	5.65	0.90	.60	.43	(.77)							
4. Backup behaviors	5.20	0.95	.35	.45	.43	(.79)						
5. Coordination	5.53	0.84	.52	.54	.56	.56	(.80)					
6. Conflict management	5.69	0.82	.49	.55	.45	.26	.45	(.77)				
7. Motivating others	5.56	0.81	.29	.30	.37	.31	.27	.45	(.86)			
8. Affect management	6.18	0.71	.36	.25	.28	.05	.22	.39	.22	(.85)		
9. Extraversion	3.15	0.65	.23	.13	.20	.18	.06	.28	.22	.02	(.86)	
10. Agreeableness	3.82	0.77	.20	.12	.10	.22	.10	.07	.17	.06	.36	(.88)
11. Cognitive abilities	104.40	16.54	−.09	−.06	−.14	.05	−.02	.09	.09	−.07	.08	.01

Note. $N = 176$. Correlations above .17 are statistically significant.

the teamwork dimensions and three individual attributes: agreeableness, extraversion, and cognitive abilities. Of the big-five personality traits, agreeableness and extraversion are the ones consistently showing positive correlations with teamwork behaviors (e.g., Morgeson, Reider, & Campion, 2005; Neumann & Wright, 1999). Confirmatory evidence would be provided if the dimensions in the resulting instrument follow a set of correlations with agreeableness and extraversion in the predicted direction (positive).

Conversely, research indicates that cognitive abilities tend to exhibit negligible or even negative correlations with behaviors associated with teamwork (e.g., Kikcul & Neuman, 2000; Neuman & Wright, 1999). As such, discriminant evidence will be supported if the behavioral factors included in the resulting measure reproduce a similar pattern of negligible (or negative) correlations.

Personality measure

We relied on Goldberg's (1992) 100-markers scale to assess agreeableness and extraversion. For purposes of the study, the 40 items (20 per trait) loading on agreeableness and extraversion were used. Items consist of simple-word adjectives and asks participants to rate the extent to which the adjective describes them. A 7-point Likert-type scale with responses ranging from 1 (*extremely accurate*) and 7 (*extremely inaccurate*) accompanied items. Example items for agreeableness include considerate, kind, and warm. Example items for extraversion include assertive, energetic, and talkative. Goldberg (1992) provided evidence on the construct validity and reliability of the 100-marker scale. Reliability scores for this study were .88 for agreeableness and .86 for extraversion.

Cognitive ability measure

Cognitive abilities were tested with the 50-question Wonderlic test. The test was administered online and requires 12 min for completion. The national average score on the test is 21.0 ($SD = 7.12$). The provider also reports a conversion of the scale with a typical standardized distribution around a mean of 100. For research purposes, we utilized the distribution with the 100-score average. Reliability and validity of the test have been supported elsewhere (e.g., McKelvie, 1994).

Table 3 shows correlations among the eight teamwork lower-level dimensions, agreeableness, extraversion, and cognitive abilities. As expected, five out of the eight correlations between extraversion and teamwork processes were positive and statistically significant ($p < .05$). And although three failed to reach significance, all eight were in the expected (positive) direction. Something similar happened with agreeableness in the sense that all correlations were positive, although only three out of the eight reached statistical significance. Cognitive ability, on the other hand, followed a distinct pattern. While none of the correlations was statistically significant, five were of negative value.

To further corroborate discriminant and convergence evidence, we created scores for each of the three higher-level dimensions—transition, action, and interpersonal—by averaging scores of their corresponding lower-level factors. As expected, extraversion exhibited positive and statistically significant ($p < .05$) correlations with all three higher-level dimensions (transition = .17, action = .18, interpersonal = .20). Agreeableness followed a similar pattern (transition = .19, action = .17, interpersonal = .17). Finally, cognitive abilities showed a set of negligible and negative correlations with the three higher-level processes (transition = −.12, action = −.13, interpersonal = −.06). One could argue the correlations between personality traits (agreeableness and extraversion) and the three higher-level dimensions are small in terms of conventional benchmarks (Cohen, 1992). Yet, our outcomes mirror the magnitude of effect sizes from meta-analytic studies that report correlations between these two traits and interpersonal actions (e.g., Chiaburu, Oh, Berry, Li, & Gardner, 2011). Together, we interpret results as convergent and discriminant evidence.

Study 2: Stability of the measure across raters

To test the stability of the configural structure in the 24-item measure, we adapted the items to peer-assessment format. Our goal was to explore the extent to which the eight-factor solution replicates data when a distinct set of raters (peers) assessed the teamwork categories. For the second study, the sample consisted of 173 graduate MBA students attending courses that require team projects. The average age of participants was 31.20 years old ($SD = 3.45$ years) with 38% being women. Mimicking the procedures of Study 1, we applied the survey toward the end of the semester. Data were collected anonymously, and participants were clearly informed via consent form that participation was voluntary and had no implications on course grading. Using a list of teams provided by the instructors, we randomly selected the target classmate to be assessed. To avoid repeated-measures bias, we ensured that there was no overlap between this sample and the one participating in Study 1.

We conducted CFA on the peer assessment data using EQS 6.1 with maximum likelihood estimation as a method for parameter calculations. Results in Table 2 (Model 4) indicate that the eight-factor solution appropriately replicated the study data (CFI = 0.914; IFI = 0.916; RMSEA = 0.072). We interpreted this outcome as preliminary evidence of the stability of the factor-structure underlying the measure across raters. To further test the factorial invariance of the instrument, we followed Byrne's (2006) recommendations. Specifically, Byrne advised for a set of sequential steps with increasing levels of restricting hypotheses in which multigroup data (self and peer assessments) are simultaneously tested.

The first step consists of establishing baseline models for each sample. Given that the instrument can operate differently across samples, Byrne warns that the baseline models may not be completely identical. To establishing baseline models, we sought for LMT modifications that could significantly improve model fit. The LMT statistics associated with Model 3 in Table 2 indicate that the suggested modifications provide marginal improvements. As such, a decision was made to use Model 3 as the baseline model for the self-assessment data. On the other hand, LMT statistics associated with Model 4 (peer data) indicate that adding a correlation between the errors of items 19 and 20 statistically enhances model fit. Model 5 in Table 2 shows fit statistics when this parameter was added (CFI = 0.924; IFI = 0.925, RMSEA = 0.068). As such, Model 3 (self-assessment) and Model 5 (peer assessment) were used as baseline models.

The next step consists of a multigroup test in which both baseline models are simultaneously inputted. This step tests the stability of the configural model across samples by exploring if the eight-dimension hierarchical structure replicates in both data sets. EQS provides a single set of fit statistics testing the suitability of the data in replicating the parameters across models. In cases of appropriate fit, chi square exhibits a value close to the addition of the independent models. Results in Table 2 (Model 6) indicate a proper model fit (CFI = 0.924; IFI = 0.926; RMESA = 0.064). Also, chi square shows an outcome (818.83) that closely replicates the sum of the independent models (Model 3 χ^2 + Model 5 χ^2 = 815.39).

In the final step, we imposed equality constraints to the multigroup model (Model 6). Because we were interested in testing the invariance of the measurement model, we constrained the freely estimated item-to-factor loadings to be similar across models. That is, we tested if items' loading performed similarly across samples. In this case, marginal deterioration of fit statistics provides evidence of the invariance of the measurement model (Byrne, 2006). Results associated with Model 7 in Table 2 show adequate fit statistics (CFI = 0.916; IFI = 0.917; RMSEA = 0.066) with marginal deteriorations when compared with Model 6. Additionally, a direct comparison to Model 6 via chi-square analysis indicates no statistical difference in model fit, $\Delta\chi^2(17)$ 836.84 − 818.83 = 18.01, $p > .05$. We interpret this result as an indication of measurement invariance across raters.

Discussion

Years ago, O'Neil, Allred, and Baker (1997) called for efforts to address the overdue task of developing adequate teamwork assessments. While significant progress has been achieved, it is our position that existing assessments still need refining, especially considering emerging evidence of the dimensionality of teamwork. In this article, we report on the development of a measure that captures an empirically-grounded factor structure of teamwork and two studies devoted to testing its psychometric properties. Results indicate that the resulting measure exhibits adequate internal-factor structure and appropriate psychometrics across raters. In the following paragraphs we discuss results of both studies in the context of conceptual analysis of teamwork modeling. We also provide guidelines for applying the measure in education.

Models of teamwork

Results with respect to the underlying structure of teamwork merit attention. We started with the notion that ten factors, clustered around three higher-level dimensions, properly represent the factor structure of teamwork. Yet, results of both studies pointed at an eight-factor solution as a superior representation. Challenging the original Mark et al. (2001) proposition was not our intent. Yet, testing the tenability of this model was a necessary condition for the development of the assessment, especially given the scarcity of studies reporting on the entire hierarchical structure. We view outcomes associated with the eight-factor solution as a worthy study contribution. Not only did Cannon-Bowers and Bowers (2011) suggested that the eight-factor solution is an alternative conceptualization, but also Mark et al. admitted that the 10-factor solution should be viewed as a guideline, thus inviting researchers to explore alternative compositions of teamwork. As Mark et al. stated, the 10-factor model is "intended to shape future conceptualization of both the scope and boundaries of team processes" (p. 370).

One could argue that LePine et al.'s (2008) meta-analysis provides empirical evidence in support for Mark et al.'s (2001) model and, as such, our results might be at odds with their findings. Yet, LePine et al.'s study primarily concentrated on exploring the role and definition of the three higher-order dimensions, something our results are not challenging. In fact, LePine et al. overlooked tests of distinct lower-order compositions despite reporting a .61 average correlation among the lower-order dimension, a statistic that suggests overlaps. Furthermore, Le Pine et al.'s results coincide with our study in the sense that the between-factor correlations within the three transition processes and within the four action processes are the ones exhibiting the highest correlations (.77 and .68, respectively). These outcomes reinforce Cannon-Bowers and Bowers's (2011) call for simplifying the factorial composition of transition and action processes. We expect future research to consider the eight-factor solution emerging from our study as a viable alternative to represent teamwork behaviors and as criteria to assess results of educational efforts.

Future researchers should also explore the stability of the eight-factor solution of teamwork and, in doing so, add evidence on the psychometric properties of the measure in the Appendix. To ensure the eight-factor solution is not sample or context specific, future researchers should consider distinct samples (e.g., employees), raters (e.g., supervisors), and settings (e.g., organization). The external validity of research conducted in academia has been questioned since McNemar (1946) described it as the science of sophomore.. While still an ongoing debate (e.g., see Peterson & Merunka, 2014), a plausible concern with respect to our findings deals with the interaction between the setting and the behaviors under scrutiny. Do class projects demand the same type of team interactions than organization projects? Generalization of the eight-factor solution warrants replication in organization settings. The primary goal of our study was to develop of an instrument to assess teamwork in academia, but the validity of the instrument—especially with respect to exploring the employability of higher education graduates—rests on the generalizability of the underlying factor structure.

Use of the measure

Most of the panel experts who provided feedback on the survey praised its transparent format. We reiterate the importance of using this format, which can be implemented by adding a definition of each dimension on top of the corresponding set of items. Not only do definitions assist in interpreting the intent of items but also clearly indicate to raters the boundaries of dimensions. We suggest using the original Mark et al. (2001) definitions as guidelines when adding these notes.

The resulting survey can be instrumental to target team-level phenomena. Although originally developed to assess individual behaviors, appraisals that target collective processes can also benefit from the resulting measure. Consistent with composition theories, aggregation indices (e.g., interclass correlation coefficient) can assist in testing the validity of studies interested in team-level processes.

Similarly, the resulting measure may be adapted across raters. As the results of Study 2 indicate, ratings distinct from self-assessments can be obtained with no significant detriment to the configural model of the measure. We view the measure, as shown in Appendix A, as appropriate for instructors interested in multisource feedback.

Conclusions

Teamwork is a cornerstone in the functioning of organizations and consequently, a key competence for the employability of higher education graduates (e.g., Betta, 2016; Robbins, 1994). Over the last decade, teamwork has been one of the top three skills employers screen in recruitment and selection (National Association of Colleges and Employers, 2014). Previously, we expressed concerns with respect to methods used to validate the educational efforts devoted to advancing the competence. Such concerns led us to develop an instrument that reflects current theorizing on the domain of teamwork. Two studies indicated that the resulting instrument (Appendix A) exhibits adequate psychometric properties and stability across raters. We hope the instrument assists instructors interested in monitoring how teamwork evolves in academia, and we call for future researchers to further support the validity of the instrument.

References

Aguado, D., Rico, R., Sanchez-Manzanares, M., & Salas, E. (2014). Teamwork competency test (TWCT): A step forward on measuring teamwork competencies. *Group Dynamics: Theory, Research, and Practice*, *18*, 101–121. doi:10.1037/a0036098

Baker, D. P., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors*, *34*, 469–475. doi:10.1177/001872089203400408

Betta, M. (2016). Self and others in team-based learning: Acquiring teamwork skills for business. *Journal of Education for Business*, *91*, 69–74. doi:10.1080/08832323.2015.1122562

Bravo, R., Lucia-Palacios, L., & Martin, M. (2016). Process and outcomes in student teamwork. An empirical study in a marketing subject. *Studies in Higher Education*, *41*, 302–320. doi:10.1080/03075079.2014.926319

Byrne, B. M. (2006). *Structural equation modeling with EQS*. Mahwah, NJ: Erlbaum.

Cannon-Bowers, J. A., & Bowers, C. (2011). Team development and functioning. In S. Zedeck (Eds.), *APA handbook of industrial and organizational psychology* (Vol. 1, pp. 597–650). Washington, DC: American Psychological Association.

Cela-Ranilla, J. M., Esteve-Mon, F. M., Esteve-Gonzalez, V., & Gisbert-Cervera, M. (2014). Developing self-management and teamwork using digital games in 3D simulations. *Australian Journal of Educational Technology*, *30*, 634–651.

Center for Creative Leadership. (2015). *The state of teams*. Retrieved from http://insights.ccl.org/wp-content/uploads/2015/04/StateOfTeams.pdf.

Chiaburu, D. S., Oh, I., Berry, C. M., Li, N., & Gardner, N. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, *96*, 1140–1166. doi:10.1037/a0024004

Clark, C. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–313. doi:10.1037/1040-3590.7.3.309

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

Doane, D. P., & Seward, L. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, *19*, 1–18. doi:10.1080/10691898.2011.11889611

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69–106. doi:10.1111/j.1529-1006.2004.00018.x

Eby, L. T., Meade, A. W., Parisi, A. G., & Douthitt, S. S. (1999). The development of an individual-level teamwork expectations measure and the application of a within-group agreement statistic to assess shared expectations for teamwork. *Organizational Research Methods*, *2*, 366–394. doi:10.1177/109442819924003

European Observatory of Working Life. (2007). *Teamwork and high performance work organisation*. Retrieved from http://www.eurofound.europa.eu/observatories/eurwork/articles/working-conditions/teamwork-and-high-performance-work-organisation

Goldberg, L. R. (1992). The development of markets for the big-five structure. *Psychological Assessment*, *4*, 26–42. doi:10.1037/1040-3590.4.1.26

Hallam, G., & Campbell, D. (1997). The measurement of team performance with a standardized survey. In M. T. Brannick, E. Salas, & C. Price (Eds.), *Performance assessment and measurement* (pp. 155–171). London, UK: Erlbaum.

Hare, A. (2010). Theories of group development and categories for interaction analysis. *Small Group Research*, *41*, 106–140. doi:10.1177/1046496409359503

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*, 967–988. doi:10.1177/014920639502100509

Hobson, C. J., & Kesic, D. (2002). A behavioral framework for skill assessment and development in teamwork training. *International Journal of Management*, *19*, 147–154.

Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organizational Science*, *12*, 435–449. doi:10.1287/orsc.12.4.435.10635

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review*, *2*, 130–147. doi:10.1016/j.edurev.2007.05.002

Jöreskog, K., & Long, J. (1993). Introduction. In K. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 1–9). Newbury Park, CA: Sage.

Kemery, E. R., & Stickney, L. (2014). A multifaceted approach to teamwork assessment in an undergraduate business program. *Journal of Management Education*, *38*, 462–479. doi:10.1177/1052562913504762

Kikcul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAS. *Journal of Business and Psychology*, *15*, 27–51. doi:10.1023/A:1007714801558

Lerner, S., Magrane, D., & Friedman, E. (2009). Teaching teamwork in medical education. *Mount Sinai Journal of Medicine*, *76*, 318–329. doi:10.1002/msj.20129

LePine, J. A., Piccolo, R. F., Jackson, C., Mathieu, J. E., & Saul, J. (2008). A meta-analysis of teamwork processes: Test of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*, 273–307. doi:10.1111/j.1744-6570.2008.00114.x

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.

Mathieu, J. E., Gilson, L. L., & Ruddy, T. (2006). Empowerment and team effectiveness: An empirical test of an integrated model. *Journal of Applied Psychology*, *91*, 97–108. doi:10.1037/0021-9010.91.1.97

McEwan, D., Ruissen, G., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: a systematic review and meta-analysis of controlled interventions. *PLoS One*, *12*, E0169604.
McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.
McKelvie, S. (1994). Validity and reliability findings for an experimental short form of the Wonderlic personnel test in academic settings. *Psychological Reports*, *75*, 907–910. doi:10.1371/journal.pone.0169604. doi:10.2466/pr0.1994.75.2.907
McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, *43*, 289–374. doi:10.1037/h0060985
Morgeson, F. P., Reider, M. H., & Campion, M. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, *58*, 583–611. doi:10.1111/j.1744-6570.2005.655.x
Mortensen, M., & Hinds, P. J. (2001). Conflict and shared identity in geographically distributed teams. *International Journal of Conflict Management*, *12*, 212–238. doi:10.1108/eb022856
National Association of Colleges and Employers. (2014). *Job outlook: The candidate skills/qualities employers want*. Retrieved from http://www.naceweb.org/s11122014/job-outlook-skills-qualities-employers-want.aspx
Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive abilities. *Journal of Applied Psychology*, *84*, 376–389. doi:10.1037/0021-9010.84.3.376
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
O'Neil, H. F. Jr., Allred, K., & Baker, E. (1997). Review of workforce readiness theoretical frameworks. In H. F. O'Neil Jr. (Eds.), *Workforce readiness: Competencies and assessments* (pp. 3–25). Mahwah, NJ: Erlbaum.
O'Neill, T. A., Goffin, R., & Gellatly, I. R. (2012). The knowledge, skill, and ability requirements for teamwork: Revisiting the Teamwork-KSA test's validity. *International Journal of Selection and Assessment*, *20*, 36–52. doi:10.1111/j.1468-2389.2012.00578.x
Peterson, R. A., & Merunka, D. R. (2014). Convenience samples of college students and research reproducibility. *Journal of Business Research*, *67*, 1035–1041. doi:10.1016/j.jbusres.2013.08.010
Robbins, T. (1994). Meaningfulness and community in the classroom: The role of teamwork in business education. *Journal of Education for Business*, *69*, 312–316. doi:10.1080/08832323.1994.10117706
Rousseau, V., Aube, C., & Savoie, A. (2006). Teamwork behaviors: A review and an integration of frameworks. *Small Group Research*, *37*, 540–570. doi:10.1177/1046496406293125
Salas, E., Diaz-Granados, D., Klein, C., Burke, C. S., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors*, *50*, 903–933. doi:10.1518/001872008X375009
Salas, E., Sims, D., & Burke, C. (2005). Is there a "big-five" in teamwork? *Small Group Research*, *36*, 555–599. doi:10.1177/1046496405277134
Senior, B., & Swailes, S. (2007). Inside management teams: Developing a teamwork survey instrument. *British Journal of Management*, *18*, 138–153. doi:10.1111/j.1467-8551.2006.00507.x
Stevens, M. J., & Campion, M. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*, 207–228. doi:10.1177/014920639902500205
Valls, V., Gonzalez-Roma, V., & Tomas, I. (2016). Linking educational diversity and team performance: Team communication quality and innovation team climate matter. *Journal of Occupational and Organizational Psychology*, *89*, 751–771. doi:10.1111/joop.12152
Viles, E., Zarraga-Rodriguez, M., & Jaca, C. (2013). A tool to assess teamwork performance in higher education. *Intangible Capital*, *9*, 281–304.

Appendix A

Teamwork measure

In teams, this person …

Transition Processes

Mission analysis

1. Participates in defining the major goals for the team.
2. Assists in articulating the team tasks and in planning task achievement
3. Understands what his/her role toward team-goal accomplishment is.

Strategy formulation

1. Participates in identifying key roadblocks to achieve team's goals and in finding solutions to deal with these obstacles.
2. Assists in discussing alternative course of action to achieving team's goals.
3. Considers alternative actions towards mission accomplishment.

Action processes

Situation monitoring

1. Tracks where the team stands in relation to its goals.
2. Monitors the contributions of each team member towards goal accomplishment.
3. Is aware of what it takes to achieve the team goal.

Backup behaviors

1. Provides verbal feedback to teammates and, if necessary, coach them.

2. Assists teammates in carrying out their actions.
3. Is willing to assume a task for a teammate.

Coordination

1. Collaborates in orchestrating teammates' interdependent actions.
2. Is flexible in accommodating to the team's requirements.
3. Gets involve in coordinating team members' contributions.

Interpersonal processes

Conflict management

1. Emphasizes common goals rather than dwelling on differences.
2. Finds ways to work through interpersonal disagreements among team members.
3. Gets involve when conflicts might represent a threat to team's goals.

Motivating and confidence building

1. Encourages team members to do their best and boost their confidence level.
2. Publicly praises the achievement of teammates.
3. Energizes other teammates towards achieving their tasks.

Affect management

1. Controls his/her temper and handles situations rationally.
2. Controls his/her emotions, so that the team can move toward achieving goals.
3. Contains any emotion that might be detrimental to the team.