Measuring Teamwork in Health Care Settings

A Review of Survey Instruments

Melissa A. Valentine, MPA, Ingrid M. Nembhard, PhD,† and Amy C. Edmondson, PhD**

Background: Teamwork in health care settings is widely recognized as an important factor in providing high-quality patient care. However, the behaviors that comprise effective teamwork, the organizational factors that support teamwork, and the relationship between teamwork and patient outcomes remain empirical questions in need of rigorous study.

Objective: To identify and review survey instruments used to assess dimensions of teamwork so as to facilitate high-quality research on this topic.

Research Design: We conducted a systematic review of articles published before September 2012 to identify survey instruments used to measure teamwork and to assess their conceptual content, psychometric validity, and relationships to outcomes of interest. We searched the ISI Web of Knowledge database, and identified relevant articles using the search terms *team*, *teamwork*, or *collaboration* in combination with *survey, scale, measure,* or *questionnaire*.

Results: We found 39 surveys that measured teamwork. Surveys assessed different dimensions of teamwork. The most commonly assessed dimensions were communication, coordination, and respect. Of the 39 surveys, 11 met all of the criteria for psychometric validity, and 14 showed significant relationships to nonself-report outcomes.

Conclusions: Evidence of psychometric validity is lacking for many teamwork survey instruments. However, several psychometrically valid instruments are available. Researchers aiming to advance research on teamwork in health care should consider using or adapting one of these instruments before creating a new one. Because instruments vary considerably in the behavioral processes and emergent states of teamwork that they capture, researchers must carefully evaluate the conceptual consistency between instrument, research question, and context.

Key Words: teamwork, psychometric properties, survey instruments

(*Med Care* 2013;00: 000–000)

From the *Harvard Business School, Boston, MA; and †Yale University School of Public Health and School of Management, New Haven, CT.

The authors declare no conflict of interest.

Reprints: Melissa A. Valentine, MPA, Harvard Business School, Boston, MA 02163. E-mail: mvalentine@hbs.edu.

ISSN: 0025-7079/13/000-000

The use of teams has grown significantly in health care organizations, becoming a critical part of the way in which care is delivered.[1,2] To deliver quality care, diverse professionals with unique expertise must often work together, such that team*work* is an essential aspect of health care delivery, regardless of whether health professionals are assigned to designated teams.[3–5] The benefits of effective teamwork can be substantial. Higher team functioning is associated with better patient outcomes[6–8] and cost savings.[9] Scholars have theorized that these benefits accrue because better functioning teams make better quality decisions, cope more effectively with complex tasks, and better coordinate actions and expertise.[9–11]

Despite growing awareness of potential benefits, many heath care organizations lack effective teamwork, with negative consequences for patients.[1] In a review of 54 malpractice incidents in an emergency department, 8 of 12 deaths were judged to have been preventable if appropriate teamwork had occurred.[12] The prevalence of teamwork failures has been attributed to several factors. First, the professional hierarchy in medicine inhibits teamwork because both high-status and low-status individuals may avoid open conversation for fear of embarrassment or disrupting the hierarchy.[13–15] Second, frequent transitions between caregivers because of shift changes, patient transfers, and academic hospital schedule constraints make coordination and teamwork complicated.[16] Finally, teamwork confronts the challenges of managing human relationships and personalities.[17]

In sum, prior research indicates that teamwork promotes quality care, worker satisfaction, and cost improvement but may not happen naturally for a number of reasons. Given its importance in health care, systematic empirical study is needed to better understand the behaviors that comprise teamwork, the factors that support teamwork, and the relationships between teamwork and health care outcomes. Such study depends in part on access to appropriate measures of teamwork. However, a systematic review of the conceptual and psychometric properties of available survey instruments has not yet been conducted.

In this paper, we report the results of our review of surveys examining teamwork. Our aim was to assist with survey selection by providing a comprehensive review of the dimensions of teamwork assessed by each survey along with the psychometric validity of the measures. We begin by reviewing what is meant by teamwork and how the meaning varies by context.

Reviews of teamwork research show that there is no consensus on what constitutes teamwork.[18–20] The literature

includes various conceptual models and multiple dimensions of teamwork (eg, Hackman's model of team effectiveness[21] and Salas' "Big Five" teamwork model[22]). Despite differences between models, "teamwork" generally refers to behavioral processes that people use to accomplish interdependent work, and/or the affective, cognitive, and motivation states that emerge during the course of that work.[20] Behavioral processes include actions such as communication, coordination, sharing expertise, and helping. Emergent states include, for example, mutual respect and psychological safety.

In this paper, we do not promote a single conceptual model, adjudicate differences between models, or propose a new theory of teamwork. Instead, we focus on how teamwork is measured in the existing research literature so as to facilitate both conceptual and empirical progress on the topic. Not all team contexts are alike, and the first decision a researcher faces is: what measures are best suited for the research setting and type of team. In some settings, tasks are performed by teams that are bounded, meaning their membership is clearly defined and is known to team members and team stakeholders.[21] In other settings, tasks are performed by large workgroups with shifting members (eg, MDs and nurses working in intensive care units) such that survey items about "the team" are illogical and certain behavioral processes or emergent states may not be relevant.[23] Our review is intended to assist researchers in selecting a survey that both fits their context and has appropriate psychometric properties.

METHODS

We conducted a systematic review of medical and management research literatures to identify articles reporting the development or use of a survey instrument that measures teamwork. We began with a broad search of the ISI Web of Knowledge article database using the keywords *team, teamwork,* and *collaboration*, in combination with *survey, scale, measure,* or *questionnaire*. In addition to ISI, we searched the references of 5 highly cited literature reviews on teams.[4,24–27] We examined every referenced article to determine whether the authors used surveys to measure teamwork and also examined the references from all of the articles identified using the above 2 strategies (ISI search and review articles) to find additional teamwork surveys. This search strategy was consistent with suggestions presented in a recent report on best practices in systematic reviews of measurement properties.[28]

In total, we examined over 2100 articles in management, social science, medicine, and health services research journals. We excluded the vast majority of these articles because they were not published in peer-reviewed journals, did not empirically assess teamwork, or reported on studies that used alternate methods, such as interviews,[29,30] direct observation,[31] video analysis,[32] or behavioral marker systems.[33–35] We also excluded surveys that used an individual level of analysis,[36–39] that measured development over time,[40] or did not measure behavior.[41]

We retained 39 articles for further review. All of these peer-reviewed articles reported the development or use of a survey measuring teamwork. We reviewed each of these surveys in 2 ways. First, we reviewed the dimensions of teamwork assessed by the surveys. We then assessed the psychometric strength of each survey and whether the survey measures had demonstrated relationships with at least 1 nonself-report outcome.

Reviewing the Dimensions of Teamwork Assessed

Because the dimensions assessed by each survey necessarily relate to the developers' research purpose and the type of team studied, we divided surveys by research purpose and team type and then qualitatively assessed the dimensions of teamwork contained. We distinguished between surveys developed for the purpose of creating models of team effectiveness versus those developed for other purposes. We made this distinction because we expected that surveys focused on creating models of team effectiveness would assess specific behaviors pertinent to the theoretical models the studies were promoting, as opposed to more general teamwork behaviors that we expected to see in the other studies. All of the surveys developed to test models of team effectiveness were developed for formally organized and bounded teams.

We then divided the other surveys by the type of team described (ie, bounded teams vs. larger, unbounded workgroups such as units or departments). For each group of surveys, the first 2 authors and a research assistant independently reviewed each item in every survey and categorized each as a behavioral process, an emergent state, or other. This categorization corresponds to Ilgen et al's[20] widely cited theory of teamwork components. We then further categorized each item using the subcategories of behavioral processes and emergent states that emerged during our review. The subcategories were based on the actual language used in the survey items. In many cases there was consistency in terminology across surveys. For example, an item in Schroder et al's[42] survey was "In this practice, team members treat each other with respect" and an item in Sorra and Nieva's[43] survey was "In this unit, people treat each other with respect;" both were categorized as "respect." When the language differed, but the latent construct being described was the same, we adopted the more frequently used language or language consistent with major teamwork theories.[20,21,25,44] There was 81% agreement between authors and research assistant on the codes during the first iteration of coding and complete agreement following discussion to reconcile differences in codes.

Assessing the Psychometric Strength of Surveys and Survey Relationship to Outcomes

To assess the psychometric strength of each survey, we performed a comprehensive review of the survey's performance with respect to criteria that assess either survey reliability or validity, which are regarded as fundamental elements of a high-quality survey.[44–47] A *reliable* survey measures a construct consistently across various conditions, whereas a *valid* survey measures what it is supposed to measure. Although threshold values for the criteria we assess are well established, we note that what is ultimately

acceptable depends on research setting and purpose.[48,49] At a minimum, a good survey performs well with respect to the following criteria:

1. *Internal consistency*. Internal consistency, a component of reliability, refers to the correlation between items in a survey measure. A strong correlation between measure items suggests that items within the measure capture the same latent construct. A commonly used statistic for assessing internal consistency is Cronbach α, which ranges between negative infinity and 1.[50] In applied settings where decisions are to be made on the basis of scores, experts note that a value of 0.9 is "the minimum that should be tolerated."[51(p 245)] However, for early-stage research and newly developed surveys, a minimum value of 0.7 is generally considered to be acceptable. It indicates moderate consistency between items (70% of variance is true score variance; 30% is random measurement error variance).[48,51] Cronbach α should be used and interpreted with caution. Scales that violate assumptions of the test can result in high α scores. Moreover, the use of this statistic comes with trade-offs. First, there is a trade-off between the length of the scale (the number of items) and the desired reliability. Longer scales typically result in higher α's but may be unwieldy in certain research settings. Second, there is also a trade-off between the desire to broadly assess different dimensions within the latent construct of interest and to exhaustively assess a single dimension of the construct. For example, a measure could include 3 items that are largely synonymous, resulting in a high α, but miss important dimensions that would have also correlated with the other items, albeit with a moderately lower α score. Internal consistency can also be assessed using Raykov coefficient or a goodness-of-fit statistic.
2. *Interrater agreement(IRA) and reliability*. A good survey will elicit similar responses about the phenomenon of interest (eg, teamwork) from different judges (eg, each person in the team). Both IRA and interrater reliability (IRR) assess the level of similarity between responses provided by different judges, an indicator of survey reliability. However, these measures define similarity differently. IRA focuses on absolute consensus between judges, whereas IRR focuses on relative consistency between judges.[49,52] Both are accepted approaches for assessing similarity. IRA is traditionally assessed by the r_{wg} index, which ranges between 0 and 1 and compares the observed response variance to the variance expected, given a uniformly distributed error.[53] A r_{wg} value of 0.7 is often cited as the minimum acceptable value, although not universally.[48] The most commonly used metrics for evaluating IRR are the intraclass correlation coefficient (ICC) and the Pearson product-moment correlation, although the former has become more accepted. Although ICC is generally treated as an indicator of IRR, by method of calculation, it also assesses IRA and therefore serves as a metric for both criteria.[49] ICC values >0 indicate similarity.[54] Some have argued that, because of different foci, both IRA and IRR should be reported as standard practice.[55] Note that IRA and IRR are particularly important for surveys measuring phenomena such as teamwork that exist at the group rather than individual level. These metrics justify the aggregation of scores to the group level. When a single group is assessed, only IRA must be satisfied to justify aggregation. When multiple groups are assessed, both IRA and IRA+IRR (eg, ICC) metrics should be used to determine whether aggregation is warranted. Results of within and between analysis, which uses an analysis of variance to test whether variation between groups is greater than variation within a group, can also be used to justify aggregation. Two other measures of scale reliability, which are infrequently reported (possibly for practical reasons because they require a survey to be administered multiple times) are IRR (assessed by surveying the same person on different occasions) and test-retest (assessed by repeated surveying over time). Both are assessed by calculating the correlation between the different survey scores, which ranges from 0 to 1, with 1 representing perfect correlation between survey responses. These additional measures of reliability are rarely reported, and we do not evaluate scales using these metrics.
3. *Structural validity.* Structural validity, which provides evidence of construct validity, refers to the extent to which the items in a scale have a high covariance structure meaning that they all move together and reflect the dimensionality of the construct as expected. It answers the question: "How many concepts does the scale measure?" If a scale claims to measure 1 concept (eg, coordination), then the evidence should indicate that it measures 1 thing. It is generally established by exploratory and confirmatory factor analyses results showing that all items in the scale belong to 1 "factor." To provide this evidence, several results from factor analyses should be reported: the number of distinct factors, the percentage of variance explained by the factor structure, the values of factor loadings (ideally, >0.40), eigenvalues (ideally, >1.0), and goodness-of-fit statistics.
4. *Content validity.* The content validity criterion requires that a survey reflect the substantive realities of the construct of interest. The "gold standard" for establishing content validity is triangulation, defined as "the combination of methodologies in the study of the same phenomenon."[56(p 602)] Researchers triangulate by comparing survey results to data obtained by observation, semistructured interviews, qualitative work, and/or expert or respondent reviews of the survey.[57,58] This comparison minimizes the risk that a survey captures a priori assumption about what is important in the construct rather than the true dimensions.

RESULTS

Each of the 39 peer-reviewed articles, we identified, reported the development or use of a survey measuring teamwork. All surveys were published during the last 20 years. More appeared in health services or medical journals (20 surveys) when compared with general management journals (19 surveys). One was published in both.[59,60]

The Dimensions of Teamwork in Surveys

Of the 39 surveys developed to measure teamwork, 9 were developed to test full models of team effectiveness. Thus, other variables in the proposed model—organizational context, team design, task design, and team performance—were assessed along with teamwork (Table 1).

Of the remaining 30 surveys, 16 were used to assess teamwork in bounded teams, and 14 assessed teamwork in unbounded workgroups in which teamwork mattered. For the 16 surveys that focused on bounded teams, the most commonly assessed behavioral dimensions of teamwork were communication and coordination, and the most commonly assessed emergent states were respect and group cohesion (Table 2).

Of the 14 surveys that examined teamwork in larger, unbounded groups, 12 focused on behavioral processes and emergent states and 2[84,85] focused on attitude towards teamwork. The 12 surveys of larger, unbounded groups were all developed in health care settings (Table 3). The behavioral dimensions that were most frequently assessed were communication and use of all contributors' expertise. The emergent states most commonly assessed were respect and social support.

On average, surveys developed for unbounded workgroups assessed more dimensions of teamwork than those developed for bounded teams. However, they did not assess group cohesion, which was commonly assessed in bounded teams. Across both team types, there was more focus on behavioral processes than on emergent states.

The Psychometric Validity of Teamwork Surveys

Only 16 of the 39 teamwork surveys (41%) were reported with all 4 psychometric properties evaluated in this review; 11 satisfied the minimum standards for all 4 criteria (Table 4). Those that completely satisfied the minimum standards are indicated by an "X" in a shaded square in the first column of Table 4. The surveys that reported all of the psychometric properties but did not satisfy all of the criteria, typically missed a cut-off point by a narrow margin (eg, Shortell et al[8] reported an α value of 0.64, which is just below the threshold of 0.70).

Of the 23 that did not report values for all of the psychometric properties evaluated, 22 did not report IRA or IRR and 1 did not report structural validity.[59]

The Relationship between Surveys and Outcomes of Interest

Of the 39 teamwork surveys identified, 14 had documented relationships with nonself-reported outcomes: 6 with clinical outcomes,[43,60,77,81,84,88] 6 with a nonclinical performance metric,[16,62,63,65,78,92] and 2 with both clinical and nonclinical outcomes.[8,72] These are indicated in column 2 of Tables 1–3. Of the remaining 22 surveys, 9 had not been examined relative to an outcome (ie, the article only reported the development of the survey) and 13 had been examined for relationships with self-reported outcomes or proposed antecedents of teamwork (eg, organizational culture). Notably, the 13 surveys with a documented relationship to a nonself-reported outcome were more likely to be reported with all 4 psychometric properties (columns 1 and 2 in Tables 1–3).

DISCUSSION

Teamwork has been an active area of research because of its potential importance in quality improvement, health

TABLE 1. Teamwork Dimensions Assessed in Surveys Testing Full Models of Team Effectiveness

	Pinto et al[61]	Campion et al[62]	Vinokur-Kaplan[63]	Denison et al[64]	Edmondson[65]	Bateman et al[66]	Doolen et al[67]	Wageman et al[16]	Senior and Swailes[68]
INPUTS									
Organizational context	X	X	X	X	X		X	X	X
Team design	X	X	X	X	X	X	X	X	X
Team task design		X	X	X			X	X	X
MEDIATORS (teamwork)									
Behavioral processes	Cooperation	Workload sharing Communication	Effort Use of expertise Strategy	Workload sharing Use of expertise Strategy	Team learning	Use of resources	Information sharing Team processes	Effort Use of expertise Strategy Social interactions	Task interactions
Emergent states		Social support Potency		Norms Teamwork Values	Psychological safety Team efficacy	Team synergy			Social support
OUTPUTS	X	X	X	X	X	X	X		
Psychometric validity					X	X	X		
Related to outcomes		X	X		X			X	

Surveys listed in columns, sorted chronologically. Team effectiveness dimensions listed in rows, sorted by Input-Mediator-Output categories.[20] An X in the second to bottom row (psychometric validity) indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the bottom row (related to outcomes) indicates that a survey has an established relationship with a nonself-reported outcome.

TABLE 2. Dimensions of Teamwork Assessed by Surveys Developed for Bounded Teams

	Brannick et al[69]	Seers[70]	Kahn and McDonough[71]	Anderson and West[72]	Hauptman and Hirji[73]	Millward and Jeffries[74]	Pearce and Sims[75]	Strasser et al[76]	Alexander et al[77]	Hoegl and Gemuenden[78]	Friesen et al[79]	La Duckers et al[80]	de Wet et al[81]	Cooper et al[82]	Strasser et al[83]	Schroder et al[42]
Behavioral processes																
General teamwork quality			X				X						X		X	
Communication	X	X		X	X	X	X	X	X	X		X	X	X		X
Coordination (mutual adjustment)	X	X		X	X	X				X		X		X	X	
Collaboration	X		X	X	X			X		X						X
Use of all members' expertise				X	X	X		X		X					X	X
Help each other/ share workload		X		X		X				X						
Shared decision making				X					X	X						
Active conflict management								X		X			X		X	X
Effort										X						
Emergent states																
Affective																
Respect				X		X		X	X	X	X		X			X
Group cohesion/ shared identity	X			X					X	X	X					
Social support				X						X	X		X			X
Psychological safety				X				X	X				X			X
Cognitive																
Role responsibility understanding		X				X		X				X				X
Shared objectives			X	X		X		X								
Psychometric validity				X		X	X		X	X				X		
Related to outcomes				X					X	X					X	

Surveys listed in columns, sorted chronologically. Teamwork dimensions listed in rows. An X in the second to bottom row (psychometric validity) indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the bottom row (related to outcomes) indicates that a survey has an established relationship with a nonself-reported outcome.

TABLE 3. Dimensions of Teamwork Assessed by Surveys Used for Larger Workgroups

	Shortell et al[8]	Baggs[84]	Adams et al[85]	Gittell[59]	Sorra and Nieva[43]	Copnell et al[86]	Hutchinson et al[87]	Sexton et al[88]	van Beuzekom et al[89]	Masse et al[90]	Ushiro[91]	Kalisch et al[92]
Behavioral processes												
General teamwork quality			X				X		X			
Communication	X	X	X	X	X	X	X	X		X	X	X
Use of all contributors' expertise	X	X	X	X		X		X		X	X	X
Coordination (mutual adjustment)	X	X	X	X	X	X		X			X	X
Collaboration		X	X		X	X		X			X	X
Active conflict management	X		X	X				X		X	X	X
Effort	X		X								X	X
Shared decision making	X	X	X			X						
Help each other/ share workload			X									X
Emergent states												
Affective												
Respect	X		X	X	X			X		X	X	X
Social support			X		X			X			X	X
Psychological safety					X			X		X		
Cognitive												
Role responsibility understanding			X	X					X		X	X
Shared objectives				X							X	X
Psychometric validity	X			X								X
Related to outcomes	X	X		X	X			X				X

Surveys listed in columns, sorted chronologically. Teamwork dimensions listed in rows. An X in the second to bottom row (psychometric validity) indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the bottom row (related to outcomes) indicates that a survey has an established relationship with a nonself-reported outcome.

care delivery, and patient safety. Many surveys have been developed to assess teamwork, and variation in measures of teamwork can give rise to confusion for health researchers. Some variation reflects differences in research focus (ie, developing a model of team effectiveness vs. testing specific antecedents of teamwork) or team type (ie, bounded or unbounded). Amidst this variety, measures of communication, coordination, respect, and use of members' expertise consistently appeared in studies, even with their different foci and team types. This consistency suggests that these are viewed as core dimensions of teamwork.

We also found variation in the quality of measures. Only 11 of the 39 surveys satisfied standard psychometric criteria, and only 5 of those showed significant statistical associations with nonself-reported outcomes. A few surveys missed the cut-offs values by narrow margins. The majority of the surveys failed to either meet or report the standard psychometric criteria expected of survey instruments. Evidence for 2 of the 4 criteria—IRA and content validity—were rarely reported. Both, along with internal consistency and structural validity, are critical to establishing statistical validity and reliability. IRA demonstrates how well a measure gathers reliable information, and content validity is important for assessing whether it captures substantive reality.[57] Without this information, others cannot evaluate the appropriateness of surveys or of individual measures for their own use. At least 1, and possibly more, indicator of each of the 4 established psychometric criteria should be reported as standard practice. Researchers, editors, and reviewers can help this become standard practice by encouraging colleagues to report surveys' complete psychometric properties.

As noted, of the 39 teamwork surveys identified, only 14 had a demonstrated relationship to nonself-reported outcomes. In addition, in our review, we observed a tendency to develop new surveys for projects, rather than adapt existing surveys, which limits the production of cumulative knowledge. The field may benefit from the use of existing, psychometrically valid surveys across studies. Consistent with this idea, 1 survey instrument—the Hospital Survey on Patient Safety Culture—deserves special note because it is being used in a national effort in the United States to collect comparative patient safety culture data from hospitals.[43] The Agency for Healthcare Research and Quality collects responses to this survey from a broad sample of hospitals and

TABLE 4. Psychometric Properties of Survey Instruments That Measure Teamwork

X	Scale	Source	No. Items, Dimensions of Teamwork Assessed, Response Scale	Interrater Agreement and Reliability*	Internal Consistency†	Content Validity	Structural Validity	Validated Relationships to Outcomes of Interest
SURVEYS TESTING FULL MODELS OF TEAM EFFECTIVENESS (TABLE 1)								
	Cross-functional Cooperation	Pinto et al[61]	Cross-functional cooperation scale, 15 items Full model 7-point Likert scale	Not reported	Cross-functional Cooperation scale, 0.92	Items informed by formal pretests, questionnaires, and follow-up interviews	Not reported	Positively associated with self-report task project outcomes
	Work Group Effectiveness	Campion et al[62]	Full survey, 54 items, 3 items each in communication/ cooperation within work group, participation Full model 5-point Likert scale	Full survey, 0.50–0.87 Communication/ cooperation scale 0.80 Participation scale 0.66	Full survey, 0.47–0.90 Communication/ cooperation scale 0.81 Participation scale 0.88	Literature review to develop items. Triangulation: team characteristics obtained from employees and managers, effectiveness obtained from employees, managers, and records	PCA confirmed that 17 of 19 team characteristics were distinct factors VarExp: 73%	Positively associated with manager perception of team effectiveness (office workers performing interdependent work) (Campion et al[62])
	Group Effectiveness Interdisciplinary Collaboration	Vinokur-Kaplan [63]/ Amer and Thomas[95]	Collaboration scale, 10 items Full model 7-point scale	Not reported	Collaboration scale, 0.82	On the basis of previously validated and implemented survey (Armer, 1978)	Not reported	Positively associated with objective standards of quality met, team cohesion, and overall team effectiveness (Vinokur-Kaplan[63])
	Team Process Domain	Denison et al[64]	Team process, 21 items Full model Scale not reported	Not reported	Team process, 0.69–0.86	Framework developed from individual and group interviews, written descriptions and team observations. Extensive testing and revision	Factor analysis suggested a 7 factor solution. FL > 0.50 EV > 1.0	Positively associated with self-report effectiveness (Denison et al[64])
X	Psychological Safety and Team Learning	Edmondson[65]	Psychologic safety, 7 items Team learning behavior, 7 items Full model 7-point scale	Intraclass correlation coefficients (ICC): Psychologic safety, 0.39 Team learning behaviors, 0.33	Psychologic safety, 0.82 Team learning behavior, 0.78	Extensive observation and interviews to develop items, extensive pretests and revisions Triangulation: confirmatory observation and interviews of teams identified by survey results as having high and low team learning behaviors	PCA confirmed that items loaded cleanly onto the 2 hypothesized factors. FL > 0.4 EV > 1.0	Positively associated with observer rated team performance (Edmondson[65])-greater team engagement in quality improvement work (Nembhard and Edmondson[15])
X	Team Effectiveness Audit Tool	Bateman et al[66]	Full survey, 46 items Full model 5-point Likert scale	Full survey, 0.97–0.98	Full survey, 0.98	Pilot questionnaire revealed themes that were used to create survey tool, which was tested and revised.	2 types of factor analysis (Cattell scree test and eigenvalues >1) identified a 4-factor solution FL > 0.3 VarExp: 72.3%	Original paper develops and validates survey instrument
X	Team Process	Doolen et al[67]	Team process, 5 items Full model 6-point Likert scale	Team processes, >0.84	Team processes, 0.818	Interviews used to qualitatively assess variables of interest. Interviews and literature review used to develop survey	Factor analysis verified team processes distinct factor ($P < 0.05$)	Positively associated with self-report team effectiveness and satisfaction (Doolen et al[67])

(*Continued*)

TABLE 4. Psychometric Properties of Survey Instruments That Measure Teamwork (*continued*)

X	Scale	Source	No. Items, Dimensions of Teamwork Assessed, Response Scale	Interrater Agreement and Reliability*	Internal Consistency†	Content Validity	Structural Validity	Validated Relationships to Outcomes of Interest
	Team Diagnostic Survey	Wageman et al[16]	Process criteria scale, 9 items Team social process, 7 items Full model 5-point Likert scale	Intraclass correlation coefficients Process criteria scale, 0.40–0.49 Team social process, 0.47	Process criteria scale, 0.89–0.92 Team social process, 0.93	Extensively validated through pretests and revisions	Comparison of within and between scale item correlation (a conservative test of structural validity) showed that the scales have weak structural validity.	Positively associated with objectively measured team performance (Wagemenet[96]) team effectiveness (Hackman and O'Connor[16])
	Team Survey	Senior and Swailes[68]	Full survey, 36 items Full model 5-point scale	Full survey, 0.68–0.90 ICC: full survey 0.38	Full survey 0.75–0.93	Repertory grid technique (described as interviews to generate constructs, analysis of constructs to generate items). Pilot test in diverse sample, tested convergent validity with Anderson and West[72]	Exploratory factor analysis found a 5-factor solution, with 1 factor including 3 subfactors. FL > 0.40 VarExp: 54%	Original paper develops and validates survey instrument
SURVEYS DEVELOPED FOR BOUNDED TEAMS (TABLE 2)								
	Team Process Scale	Brannick et al[69]	Team process scale No. items not reported 4 dimensions Response scale not reported	r_{wg} not reported; some of the scales (cooperation and giving suggestion) showed high agreement between raters, others did not	Vary widely, from 0.36–0.85 depend on rater (ie, team or observer)		Factor analysis not clearly reported; some of the scales (cooperation and giving suggestion) showed discriminant validity, others did not	Positively associated with-quality overall performance on a simulator task in the laboratory Cited in health care simulation studies of teamwork
	Team Member Exchange (TMX) Quality Scale	Seers[70]	TMX scale, 10 items 4 dimensions 5-point scale	Not reported	TMX scale, 0.83	On the basis of Seers'[70] earlier TMX scale, developed for individual level of analysis	Not reported	Gains in departmental efficiency related to average change in scale over time (Seers[70])
	Collaboration Scale	Kahn and McDonough[71]	Collaboration scale, 6 items 3 dimensions 5-point scale	Not reported	Collaboration scale, 0.92	Scale is based on literature/ previous studies	Factor analysis revealed a unidimensional construct for collaboration FL > 0.70 EV > 1 Varexp: 72%	Original study shows that collaboration is important to self-report performance and satisfaction working with other departments
	Team Climate Inventory	Anderson and West[72]	Full survey, 38 items 13 dimensions 7- or 5-point Likert scale	Full survey, 0.67–0.98	Full survey, 0.84–0.94	Literature review to develop items, extensive pretests and revisions, including pilot survey tested on sample of 155 respondents	Extensive exploratory factor analyses found 4 and 5-factor solutions with acceptable goodness-of-fit. FL > 0.5. VarExp: 62%	Positively associated with superior clinical care and patient evaluation (Bower et al[6])-patient satisfaction (Proudfoot et al[97]) -quality of work in medical laboratories (Pitt et al[98])-lower turnover in health care teams (Kivimaki et al[99])

	Team Process Quality	Hauptman and Hirji[73]	Team Process Quality Scale, 16 items, 4 dimensions 5-point ordinal scale	Not reported	Team Process Quality Scale, 0.75–0.77	Questionnaire was pretested through semistructured interviews with managers involved in new product development activities, also based on literature	FL > 0.60 EV > 1.0 VarExp: 29%	Original study shows that effective team processes overcome challenges of physical distance and time zone distance
X	Team Survey	Millward and Jeffries[74]	Full survey, 40 items 8 dimensions Unreported scale	Full survey, split half coefficient of 0.93	Full survey, 0.70–0.93	Focus group discussions and interviews with team development experts and team mangers used for revision and to develop criteria for team performance. Also adapted existing scales	Factor analysis predicted five factors, but only 4 were meaningful in psychological terms and retained. VarExp: 30%	Original paper reports significant relationship between teamwork factors and team effectiveness by an independent rater team effectiveness is not defined
X	Team Effectiveness	Pearce and Sims[75]	Team Effectiveness, 26 items 3 dimensions 5-point Likert scale	Team effectiveness, 0.85	Team effectiveness, 0.85	Measures were developed based on existing research. Team effectiveness research was based on Ancona and Caldwell,[100] Manz and Sims,[101] and Cox[102]	Factor analysis revealed a unidimensional construct for effectiveness	Team effectiveness is the outcome variable (vertical and shared leadership are predictive of greater team effectiveness)
	Team Functioning	Strasser et al[76]	Team relations, 45 items team actions 27 items true/false 8 dimensions 7-point Likert, and 10-point scale	Not reported	Team relations, 0.59–0.84 Team actions 0.73–0.93	Questions were taken from previous work and adapted for rehabilitation teams	Not reported	Original paper uses team functioning scales as an outcome variable (tested for a relationship with culture)
X	Cross-Functional Team Processes	Alexander et al[77]	Team participation, 7 items Team functioning, 8 items 7 dimensions 7-point scale (agree-disagree)	Team participation, 0.90 Team functioning, 0.88	Team participation, 0.90 Team functioning, 0.91	Based on previously validated scale	PCA confirmed 2 distinct factors as hypothesized	Team participation associated with improvements in patient functioning Team functioning was not significantly associated with patient functioning: (Alexander et al[77])
X	Teamwork Quality Survey	Hoegl and Gemuenden[78]	Teamwork scale, 37 items 13 dimensions 5-point scale	Teamwork scale, 0.79–0.95	Teamwork scale, 0.72–0.97	Literature review to develop items, pilot tests and revisions of items and structure	PCA confirmed that teamwork items loaded cleanly onto 1 factor, as hypothesized VarExp: 71.5%	Positively associated with manager-rated and team leader-rated effectiveness and efficiency (in innovative software team projects) (Hoegl and Gemuenden[78])
	Teamwork Scale	Friesen et al[79]	Teamwork scale, 9 items 3 dimensions 5-point scale	Not reported	Teamwork scale, 0.89	Focus groups used to generate constructs, which were translated into questions that were tested with a pilot group	Factor analysis supported single factor solution for teamwork scale, FL > 0.4 EV > 1 VarExp: 31%	Self-reported relationship with perceived stress (Friesen et al[79])
	Team Organization	La Duckers et al[80]	Team organization, 5 items 3 dimensions 7-point Likert scale	Not reported	Team organization, 0.84	Development included 2 phases: first a literature review and expert assessment of the clarity, completeness of questions; and pilot test to determine psychometrics	PCA revealed 3 factors FL > 0.5 VarExp: 15%	Original paper develops and validates survey instrument

(*Continued*)

TABLE 4. Psychometric Properties of Survey Instruments That Measure Teamwork (*continued*)

X	Scale	Source	No. Items, Dimensions of Teamwork Assessed, Response Scale	Interrater Agreement and Reliability*	Internal Consistency†	Content Validity	Structural Validity	Validated Relationships to Outcomes of Interest
	Primary Care Patient Safety Climate Measure (PC-SafeQuest)	de Wet et al[81]	Full survey, 30 items communication, 5 items, teamwork, 7 items 6 dimensions 7-point scale	Not reported	Communication, 0.80 Teamwork, 0.89	Literature review used to develop draft questionnaire, focus group of practitioners completed a content validity index, experts revised wording of items, survey tested with 563 members of primary care teams	Factor analysis supported 5 factors that corresponded to model. FL > 0.4 Goodness-of-fit tests included comparative fit index, root mean square error of approximation, all confirmed this tool	Original paper develops and validates survey instrument Focused on primary care collaborative practice
X	Team Emergency Assessment Measure (TEAM)	Cooper et al[82]	TEAM, 11 items 2 dimensions 5-point scale	IRR: Cohen κ, 0.5 TEAM, ICC, 0.60	TEAM, 0.98	Literature review to develop items, review of draft by experts, pilot tests on simulated and video-taped events. Focused on emergency medical teams	PCA revealed 1 factor FL > 0.6 EV > 1 VarExp: 81%	Original paper develops and validated survey instrument
	Team Functioning Survey	Strasser et al[83]	Team functioning survey, 60 items Team basics communication, 11 items 4 dimensions 7-point Likert scale	Not reported	Team basics/ communication, 0.95	Item response theory used to revise a survey previously used (development not reported)	Factor analysis supported theoretical model, although team basics and coordination were not cleanly discriminant. Team basics/ communication EV > 1, VarExp: 68%	Measures of team functioning (full scale) were associated with patient outcomes; team basics/communication associated with community discharge
	Collaborative Practice Assessment Tool (CPAT)	Schroder et al[42]	Fully survey, 56 items, relationships, 5 items, communication, 5 items, decision making, 5 items 8 dimensions 7-point Likert scale	Not reported	Relationships, 0.81, communication, 0.74, decision making, 0.74,	Literature review and use of expert team to develop items, 2 separate pilot tests and revisions to develop final instrument	CFA revealed 8 factors that matched model of collaborative practice. Tests of model fit included normed fit index, comparative fit index, Tucker Lewis index, which all reported acceptable scores	Original paper develops and validated survey instrument
SURVEYS USED FOR LARGER WORKGROUPS (TABLE 3)								
	ICU Nurse-Physician Collaboration	Shortell et al[8]	Full survey, 82 items Coordination scale, 13 items Communication scale, 43 items Problem-solving scale, 14 items 7 dimensions 5-point Likert scale	Tested using ANOVA: variance within the units significantly less than variance between units ($P < 0.05$)	Full survey, 0.61–0.88 Coordination Scale, 0.75–0.81 Communication Scale, 0.64–0.86 Problem-solving scale, 14 items	Literature review to develop items, pilot tests and revisions of items and structure Triangulation: on-site observational visits and semistructured interviews conducted after data collection to confirm that high, medium, and low scores correlated with actual high, medium, and low performance	PCA performed on a subset of measures, not reported for teamwork measures	Positively associated with lower risk-adjusted length of stay, lower nurse turnover, higher evaluated technical quality of care, and greater evaluated ability to meet family member needs in ICU (Shortell et al[8])-lower incidence of mortality and chronic, severe morbidity in NICU (Pollack et al[103])

	Collaboration and Satisfaction about Care Decisions	Baggs[86]	Collaboration scale, 7 items 5 dimensions 7-point scale	Not reported	Collaboration scale, 0.93	Literature review to develop items. Relevance and adequacy of measures confirmed by 12 nursing and medical experts	PCA confirmed 1 factor for collaboration. FL: 0.82–0.93 Var Exp: 75%	Positively associated with patient outcomes (Baggs[86])-nurse satisfaction with decision making (Dechairo-Marino et al[104])
	Professional Working Relationships	Adams et al[87]	Professional working relationships, 26 items 12 dimensions 4-point response scale	Not reported	Professional working relationships, 0.84–0.91	Extensively tested through qualitative work, literature review, pretest, and revisions	Maximum likelihood extraction, factor analysis produced similar factor structures that supported conceptual design FL > 0.3 EV > 1.0	Positively associated with nurses' job satisfaction (Adams et al[87])
X	Relational coordination	Gittell[59,105]	Relational coordination scale 7 items * n work groups 7 dimensions 5-point Likert scale	Cross-group differences in relational coordination tested using ANOVA ($P < 0.01$)[59] ICC=0.25[105]	Relational coordination scale, 0.8	Extensively tested through qualitative work (Gittell et al[60])	Exploratory factor analysis confirmed 1 factor solution[105] FL > 0.4 EV > 1.0	Positively associated with patient functional and pain status, mental health, (Gittell[59])-quality of care, postoperative functioning; negatively associated with postoperative pain and length of stay (Gittell et al[60]). Has also been used in bounded teams (Noel et al[106])
	Hospital Survey on Patient Safety	Sorra and Nieva[43]	Full survey, 42 items Teamwork within units scale, 4 items Organizational learning scale, 3 items Communication openness scale, 3 items 6 dimensions 5-point Likert scale	Not reported	Teamwork within units scale, 0.83 Organizational learning scale, 0.76 Communication openness scale, 0.73	Literature review and interviews with hospital staff to develop items. Part of a national effort to collect comparative longitudinal data on patient safety culture in acute care, primary care, pharmacy, and nursing facility settings	PCA yielded 14 factors FL > 0.4 EV > 1.0 VarExp: 64.5%	Positively associated with incident reporting behavior in the NICU (Snijders, 2009) Scores improved following teamwork training (Blegen et al[107]) Further validated in Sorra and Nieva[43]
	Perceptions about Interdisciplinary collaboration scale	Copnell et al[88]	Full survey, 29 items 5 dimensions 5-point Likert scale	Not reported	Not reported	Adapted from,Anderson[108] several measures changed. Piloted with nurses in one NICU to test face validity, slight revisions were made. Scale was developed for use in a pre/post intervention study	Not reported	Original study reported the pre and post results of an intervention—no significant changes in collaboration scores resulted from intervention
	Teamwork Scale	Hutchinson et al[89]	Teamwork scale, 22 items 2 dimensions 5-point Likert scale	Not reported	Teamwork scale, 0.69–0.84	Pretested with focus groups and frontline workers, selected for face validity.	Exploratory factor analysis confirmed 2 factor solution for teamwork domain. FL > 0.40 VarExp: 50%	Original paper develops and validates survey instrument
	Safety Attitudes Questionnaire	Sexton et al[90]	Full survey, 40 items teamwork climate scale, 14 items 8 dimensions 5-point Likert scale	Not reported	Full survey, Raykov coefficient: 0.90	Literature review to develop items, pilot tests and revisions of items and structure Adapted for NICU (Profit et al[109]); adapted into Turkish (Kaya et al[110])	CFA confirmed hypothesized 6 factor structure, teamwork scale, FL: 0.76–0.96	Communication and collaboration were associated with lower risk-adjusted morbidity, not associated with mortality (Davenport et al[7]) Scores improved following an intervention (Sexton et al[90])

(*Continued*)

TABLE 4. Psychometric Properties of Survey Instruments That Measure Teamwork (*continued*)

X	Scale	Source	No. Items, Dimensions of Teamwork Assessed, Response Scale	Interrater Agreement and Reliability*	Internal Consistency†	Content Validity	Structural Validity	Validated Relationships to Outcomes of Interest
	Leiden Operating Theater and Intensive Care Safety (LOTICS)	Van Beuzekom et al[91]	LOTICS, 40 items 2 dimensions 4-point Likert scale	Not reported	LOTICS, 0.75–0.88	A multidisciplinary ICU team made an inventory of all possible process failures; this inventory was reviewed by multidisciplinary board, which also identified the causes of the process failures. These were used to develop questions, which were reviewed by the supervisory board for readability and validity	Exploratory factor analysis revealed nine factors, FL > 0.4 VarExp: 48%	Original paper develops and validates survey instrument
	Collaboration Scale	Masse et al[92]	Collaboration scale, 23 items 5 dimensions 5-point Likert type response	Not reported	Collaboration scale, 0.75–0.91	Questions developed based on pre-existing conceptual models (Rosenfeld[111]) and adapted through a collaborative web-based exercise	Confirmatory factor analysis ruled out initial factor structure; a three factor solution was arrived at FL > 0.42	Original paper develops and validates survey instrument
	Nurse-Physician Collaboration	Ushiro[93]	Collaboration scale, 27 items 10 dimensions 7-point Likert scale	Not reported	Collaboration scale, 0.8–0.9	Scale was developed using a literature review, observation of nurse-physician exchanges in acute care hospitals, and key-informant interviews. Items were refined with pretest survey.	Exploratory factor analysis yielded 3 factors. The 3-factor model was confirmed by confirmatory factor analysis. FL > 0.4	Negatively related to nurses sex role attitudes (Ushiro[93])
X	Nursing Teamwork Survey	Kalisch et al[94]	Teamwork survey, 33 items 13 dimensions 5-point Likert scale	Full survey: 0.98 Full survey ICC: 0.16	Teamwork survey, 0.94 Scales, 0.74–0.85	Based on a theoretical framework (Salas et al[22]) focus groups conducted to develop items within categories. Experts reviewed each questions and suggested modifications or elimination	Exploratory factor analysis yielded 5 factors. The 5-factor model was confirmed by confirmatory factor analysis. FL > 0.4	Positively related to higher staffing levels (Kalisch et al[112])-job satisfaction (Kalisch et al[94])-missed nursing care (Kalisch[113,114])
SCALES MEASURING ATTITUDES TOWARDS TEAMWORK								
	Attitudes towards Health Care Teams	Heinemann et al[84]	Full survey, 28 items 4-point Likert scale	Not reported	Full survey, 0.72–0.87	Developed using focus groups, pilot test, and revision of ambiguous items	FL > 0.4 EV > 1.0 VarExp: 7.3%	Original paper develops and validates survey instrument
	Jefferson Scale of Attitudes toward Physician-Nurse Collaboration	Hojat et al[85]	Full survey, 20 items 4-point Likert scale	Not reported	Full survey, 0.84	No qualitative or pilot testing reported.	Factor analysis generated 4 factors. FL > 0.40	Original paper develops and validates survey instrument Scale later used as outcome variable

X—an X in column 1 indicates that the survey reports and meets all criteria for psychometric validity.

*Value reported is r_{wg} statistic unless otherwise indicated.

†Value reported is Cronbach α unless otherwise indicated.

AHRQ indicates Agency for Healthcare Research and Quality; ANOVA analysis of variance; CFA, confirmatory factor analysis; EV, eigenvalues; FL, factor loadings; ICC, intraclass correlation coefficient; ICU, intensive care unit; IRR, interrater reliability; NICU, neonatal intensive care unit; PCA, principal component analysis; VarExp, Variance Explained.

has made the responses available in the User Comparative Database Report since 2007.[115] Versions of this survey now exist for nursing facilities, physician practices, and pharmacies as well.[116]

A common limitation in review articles comes from having to delimit a search area and consequently losing other valuable information. Our review focused on surveys that assess teamwork, but we note that there is also a rich research literature specific to each dimension of teamwork assessed in these surveys (eg, communication, decision making, or conflict management). We did not include these dimensions as search terms and did not include other specific elements of teamwork in our review for conceptual precision and for practical reasons (eg, space constraints). Researchers may wish to search for survey measures of other specific concepts that are particularly relevant to their study.

A second limitation is that our review does not evaluate surveys on all properties known to be important for survey validity. For example, we did not analyze the wording of the surveys, although item wording can bias responses,[117] nor properties that had a time-varying component (eg, test-retest reliability). The former was not evaluated because an accepted standard for assessment does not exist, and latter was excluded because the rarity with which it was reported suggested that it should not be regarded in our minimum criteria. However, we encourage researchers to provide greater evidence of psychometric validity.

SUMMARY

This article intends to help researchers or practitioners who seek the optimal teamwork survey for use in their future work. We have argued that answer depends on a number of factors. First and foremost, there should be conceptual consistency between the survey selected and the theory explored in the research context. Second, researchers may need to consider whether and how to adapt an existing survey to a new setting. A theory of teamwork may look different in an intensive care unit than in a primary care clinic, and survey items may need to be changed to reflect these differences and then further validated. There is a trade-off between the generalizability and precision of a teamwork survey: the more generalizable a survey, the more applicable it will be for diverse settings. However, it might be more difficult to assess the particular processes in the causal pathway between teamwork and team performance if the survey is too general. Third, the survey should previously have satisfied the criteria for psychometric validity, which enables greater user confidence. We would advise against using a survey without adequate knowledge about its psychometric validity. Finally, users should consider administrative constraints. Surveys vary considerably in the number of items they contain (range, 6–82), and longer surveys may fatigue respondents.

In conclusion, this paper aims to assist the selection process by providing information about the dimensions of teamwork in and psychometric properties of existing teamwork surveys. Some researchers or practitioners may still need to develop a substantively new survey for their project. However, we advise the use of existing, psychometrically valid measures, when possible, to facilitate the development of cumulative knowledge about teamwork.

REFERENCES

1. IOM. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: New Academy Press; 2001.
2. JCAHO. *Sentinel events: Evaluating Cause and Planning Improvement*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 1998.
3. IOM. *Keeping Patients Safe: Transforming the Work Environment of Nurses*. Washington, DC: Institute of Medicine; 2003.
4. Lemieux-Charles L, McGuire WL. What do we know about health care team effectiveness? A review of the literature. *Med Care Res Rev*. 2006;63:263–300.
5. Schmitt MH. Collaboration improves the quality of care: methodological challenges and evidence from US health care research. *J Interprof Care*. 2001;15:47–66.
6. Bower P, Campbell S, Bojke C, et al. Team structure, team climate and the quality of care in primary care: an observational study. *Qual Saf Health Care*. 2003;12:273–279.
7. Davenport DL, Henderson WG, Mosca CL, et al. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg*. 2007;205:778–784.
8. Shortell SM, Rousseau DM, Gillies RR, et al. Organizational assessment in intensive-care units (ICUs)—construct development reliability, and validity of the ICU nurse-physician questionnaire. *Med Care*. 1991;29:709–726.
9. Grumbach K, Bodenheimer T. Can health care teams improve primary care practice? *JAMA*. 2004;291:1246–1251.
10. Dean PJ, LaVallee R, McLaughlin CP. Teams at the core of continuous learning. In: McLaughlin CP, ed. *Continuous Quality Improvement in Health Care: Theory Implementation and Applications*. Gaithersburg: Aspen Publishers; 1999:147–169.
11. Wagner EH. The role of patient care teams in chronic disease management. *Br Med J*. 2000;320:569–572.
12. Risser DT, Rice MM, Salisbury ML, et al. The potential for improved teamwork to reduce medical errors in the emergency department. *Ann Emerg Med*. 1999;34:373–383.
13. Edmondson A. Learning from mistakes is easier said than done: group and organizational influences on the detection and correction of human error. *J Appl Behav Sci*. 1996;32:5–28.
14. Lichtenstein R, Alexander JA, McCarthy JF, et al. Status differences in cross-functional teams: effects on individual member participation, job satisfaction, and intent to quit. *J Health Soc Behav*. 2004;45:322–335.
15. Nembhard IM, Edmondson AC. Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *J Organ Behav*. 2006;27:941–966.
16. Wageman R, Hackman JR, Lehman E. Team diagnostic survey: development of an instrument. *J Appl Behav Sci*. 2005;41:373–398.
17. Steiner I. *Group Process and Productivity*. New York: Academic Press Inc.; 1972.
18. Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. *Acad Manage Rev*. 2001;26:356–376.
19. Dickinson T, McIntyre R. A conceptual framework for teamwork measurement. In: Brannick MT, Salas E, Prince C, eds. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Lawrence Erlbaum; 1997:19–43.
20. Ilgen DR, Hollenbeck JR, Johnson M, et al. Teams in organizations: from input-process-output models to IMOI models. *Annu Rev Psychol*. 2005;56:517–543.
21. Hackman JR. *Leading Teams: setting the Stage for Great Performances*. Boston: Harvard Business School Press; 2002.
22. Salas E, Sims DE, Burke CS. Is there a "Big Five" in teamwork? *Small Group Res*. 2005;36:555–599.

23. Weinberg DB, Cooney-Miner D, Perloff JN, et al. Building collaborative capacity promoting interdisciplinary teamwork in the absence of formal teams. *Med Care*. 2011;49:716–723.
24. Bettenhausen KL. Five years of groups research: what we have learned and what needs to be addressed. *J Manage*. 1991;17:345–381.
25. Cohen SG, Bailey DE. What makes teams work: group effectiveness research from the shop floor to the executive suite. *J Manage*. 1997;23:239–290.
26. Guzzo RA, Dickson MW. Teams in organizations: recent research on performance and effectiveness. *Annu Rev Psychol*. 1996;47: 307–338.
27. Holland S, Gaston K, Gomes J. Critical success factors for cross-functional teamwork in new product development. *Int J Manag Rev*. 2000;2:231–259.
28. Terwee C. Consensus-based standards for the selection of health Measurement INstruments (COSMIN). 2012. Available at: http://www.cosmin.nl/cosmin_1_0.html. Accessed November 8, 2012.
29. Makowsky MJ, Schindel TJ, Rosenthal M, et al. Collaboration between pharmacists, physicians and nurse practitioners: a qualitative investigation of working relationships in the inpatient medical setting. *J Interprof Care*. 2009;23:169–184.
30. Slonski-Fowler KE, Truscott SD. General education teachers' perceptions of the prereferral intervention team process. *J Educ Psychological Cons*. 2004;15:1–39.
31. Healey AN, Olsen S, Davis R, et al. A method for measuring work interference in surgical teams. *Cogn Technol Work*. 2008;10:305–312.
32. Mackenzie CF, Xiao Y. Video techniques and data compared with observation in emergency trauma care. *Qual Saf Health Care*. 2003;12:II51–II57.
33. Mathieu JE, Heffner TS, Goodwin GF, et al. The influence of shared mental models on team process and performance. *J Appl Psychol*. 2000;85:273–283.
34. Malec JF, Torsher LC, Dunn WF, et al. The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul Healthc*. 2007;2:4–10.
35. Flowerdew L, Brown R, Vincent C, et al. Development and validation of a tool to assess emergency physicians' nontechnical skills. *Ann Emerg Med*. 2012;59:376–385.
36. MacDonald CJ, Archibald D, Trumpower D, et al. Interprofessional Collaborative Competencies Survey Attainment Survey. 2009. Available at: http://www.ennovativesolution.com/WeLearn/IPE-EN/ICCAS.Final.pdf Accessed October 22, 2012.
37. Weiss SJ, Davis HP. Validity and reliability of the collaborative practices scale. *Nurs Res*. 1985;34:299–305.
38. Ginsburg L, Castel E, Tregunno D, et al. The H-PEPSS: an instrument to measure health professionals' perceptions of patient safety competence at entry into practice. *BMJ Qual Saf*. 2012;21:676–684.
39. Patterson PD, Weaver MD, Weaver SJ, et al. Measuring teamwork and conflict among emergency medical technician Personnel. *Prehosp Emerg Care*. 2012;16:98–108.
40. Wheelan SA, Hochberger JM. Validation studies of the group development questionnaire. *Small Group Res*. 1996;27:143–170.
41. Gibson C, Zellmer-bruh M, Schwab D. Team effectiveness in multinational organizations: evaluation across contexts. *Group Organ Manag*. 2003;28:444–474.
42. Schroder C, Medves J, Paterson M, et al. Development and pilot testing of the collaborative practice assessment tool. *J Interprof Care*. 2011; 25:189–195.
43. Sorra J, Nieva V. Hospital Survey on Patient Safety Culture: Agency for Healthcare Research and Quality; Rockville, MD: 2004. Available at http://www.ahrq.gov/qual/patientsafetyculture/hospsurvindex.htm.
44. Mokkink L, Terwee C, Patrick DL, et al. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes: results of the COSMIN study. *J Clin Epidemiol*. 2010;63:737–745.
45. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
46. Litwin M. *How to Assess and Interpret Survey Psychometrics*. 2nd ed. Thousand Oaks, CA: Sage; 2003.
47. Allen MJ, Yen WM. *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press; 2002.
48. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria—what did they really say? *Organ Res Methods*. 2006;9:202–220.
49. LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. *Organ Res Methods*. 2008;11: 815–852.
50. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
51. Nunnally JC. *Psychometric Theory*. 2nd ed. New York: McGraw-Hill; 1976.
52. Bliese PD. Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In: Klein KJ, Kozlowski SWJ, eds. *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions and New Directions*. San Francisco: Jossey-Bass; 2000:349–381.
53. James LR, Demaree RG, Wolf G. Estimating within group interrater reliability with and without response bias. *J Appl Psychol*. 1984;69: 85–98.
54. Shrout PE, Fleiss JL. Intraclass correlations—uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
55. Klein K, Bliese PD, Kozlowski SWJ, et al, eds. Multilevel analytical techniques: commonalities, differences, and continuing questions. Multileveltheory, research, and methods in organizations: foundations, extensions, and new directions. Indianapolis, IN: Pfeiffer; 2000:512–553.
56. Denzin NK. *The Research Act*. 2nd ed. New York: McGraw-Hill; 1978.
57. Jick TD. Mixing qualitative and quantitative methods—triangulation in action. *Adm Sci Q*. 1979;24:602–611.
58. Edmondson AC, McManus SE. Methodological fit in management field research. *Acad Manage Rev*. 2007;32:1155–1179.
59. Gittell JH. Coordinating mechanisms in care provider groups: relational coordination as a mediator and input uncertainty as a moderator of performance effects. *Manage Sci*. 2002;48:1408–1426.
60. Gittell JH, Fairfield KM, Bierbaum B, et al. Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay—a nine-hospital study of surgical patients. *Med Care*. 2000;38:807–819.
61. Pinto MB, Pinto JK, Prescott JE. Antecedents and consequences of project team cross-functional cooperation. *Manage Sci*. 1993;39: 1281–1297.
62. Campion MA, Medsker GJ, Higgs AC. Relations between work group characteristics and effectiveness—implications for designing effective work groups. *Personnel Psychol Win*. 1993;46:823–850.
63. Vinokur-Kaplan D. Treatment teams that work (and those that don't): an application of Hackman's group effectiveness model to interdisciplinary teams in psychiatric hospitals. *J Appl Behav Sci*. 1995;31:303–327.
64. Denison DR, Hart SL, Kahn JA. From chimneys to cross-functional teams: developing and validating a diagnostic model. *Acad Manage J*. 1996;39:1005–1023.
65. Edmondson A. Psychological safety and learning behavior in work teams. *Adm Sci Q*. 1999;44:350–383.
66. Bateman B, Wilson F, Bingham D. Team effectiveness: development of an audit questionnaire. *J Manage Dev*. 2002;21(3/4):215–226.
67. Doolen TL, Hacker ME, Van Aken EM. The impact of organizational context on work team effectiveness: a study of production team. *Ieee Trans Eng Manage*. 2003;50:285–296.
68. Senior B, Swailes S. Inside management teams: developing a teamwork survey instrument. *Brit J Manage*. 2007;18:138–153.
69. Brannick MT, Roach RM, Salas E. Understanding team performance: a multimethod study. *Hum Perform*. 1993;6:287.
70. Seers A. Team-member exchange quality—a new construct for role-making research. *Organ Behav Hum Decis Process*. 1989;43:118–135.
71. Kahn KB, McDonough EF. An empirical study of the relationships among co-location, integration, performance, and satisfaction. *J Prod Innovat Manage*. 1997;14:161–178.

72. Anderson NR, West MA. Measuring climate for work group innovation: development and validation of the team climate inventory. *J Organ Behav*. 1998;19:235–258.
73. Hauptman O, Hirji KK. Managing integration and coordination in cross-functional teams: an international study of concurrent engineering product development. *R D Manage*. 1999;29:179–191.
74. Millward LJ, Jeffries N. The team survey: a tool for health care team development. *J Adv Nurs*. 2001;35:276–287.
75. Pearce CL, Sims HP Jr. Vertical versus shared leadership as predictors of the effectiveness of change management teams: an examination of aversive, directive, transactional, transformational, and empowering leader behaviors. *Group Dynamics: Theory Res Pract*. 2002;6: 172–197.
76. Strasser DC, Smits SJ, Falconer JA, et al. The influence of hospital culture on rehabilitation team functioning in VA hospitals. *J Rehabil Res Dev*. 2002;39:115–125.
77. Alexander JA, Lichtenstein R, Jinnett K, et al. Cross-functional team processes and patient functional improvement. *Health Serv Res*. 2005;40:1335–1355.
78. Hoegl M, Gemuenden HG. Teamwork quality and the success of innovative projects: a theoretical concept and empirical evidence. *Organ Sci*. 2001;12:435–449.
79. Friesen LD, Vidyarthi A, Baron RB, et al. Factors associated with intern fatigue. *J Gen Intern Med*. 2008;23:1981–1986.
80. La Duckers M, Wagner C, Groenewegen PP. Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives. *BMC Health Serv Res*. 2008;8:9.
81. de Wet C, Spence W, Mash R, et al. The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care*. 2010;19:578–584.
82. Cooper S, Cant R, Porter J. Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation*. 2010;81:446–452.
83. Strasser DC, Burridge AB, Falconer JA, et al. Measuring team process for quality improvement. *Top Stroke Rehabil*. 2010;17:282–293.
84. Heinemann GD, Schmitt MH, Farrell MP, et al. Development of an attitudes toward health care teams scale. *Eval Health Prof*. 1999;22:123–142.
85. Hojat M, Fields SK, Veloski JJ, et al. Psychometric properties of an attitude scale measuring physician-nurse collaboration. *Eval Health Prof*. 1999;22:208–220.
86. Baggs JG. Development of an instrument to measure collaboration and satisfaction about care decisions. *J Adv Nurs*. 1994;20:176–182.
87. Adams A, Bond S, Arber S. *Development and validation of scales to measure organisational features of acute hospital wards Int J Nurs Stud*. 1995;32:612–627.
88. Copnell B, Johnston L, Harrison D, et al. Doctors' and nurses' perceptions of interdisciplinary collaboration in the NICU, and the impact of a neonatal nurse practitioner model of practice. *J Clin Nurs*. 2004;13:105–113.
89. Hutchinson A, Cooper KL, Dean JE, et al. Use of a safety climate questionnaire in UK health care: factor structure, reliability and usability. *Qual Saf Health Care*. 2006;15:347–353.
90. Sexton JB, Helmreich RL, Neilands TB, et al. The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res*. 2006;6:10.
91. van Beuzekom M, Akerboom SP, Boer F. Assessing system failures in operating rooms and intensive care units. *Qual Saf Health Care*. 2007;16:45–50.
92. Masse LC, Moser RP, Stokols D, et al. Measuring collaboration and transdisciplinary integration in team science. *Am J Prev Med*. 2008;35:S151–S160.
93. Ushiro R. *Nurse-physician collaboration scale: development and psychometric testing J Adv Nurs*. 2009;65:1497–1508.
94. Kalisch BJ, Lee H, Salas E. The development and testing of the nursing teamwork survey. *Nurs Res*. 2010;59:42–50.
95. Armer B, Thomas BK. Attitudes Toward Interdisciplinary Collaboration in Pupil Personnel Services Teams. *Journal of School Psychology*. 1978;16:167–176.
96. Wageman R. How leaders foster self-managing team effectiveness: Design choices versus hands-on coaching. *Organization Science*. 2001;12:559–577.
97. Proudfoot J, Jayasinghe UW, et al. Team climate for innovation: what difference does it make in general practice? *International Journal for Quality in Health Care*. 2007;19:164–169.
98. Pitt SJ, Sands RL. Effect of staff attitudes on quality in clinical microbiology services. *British Journal of Biomedical Science*. 2002;59:69–75.
99. Kivimaki M, Vanhala A, et al. Team climate, intention to leave and turnover among hospital employees: Prospective cohort study. *BMC Health Services Research*. 2007;7:8.
100. Ancona DG, Caldwell DF. Bridging the Boundary - External Activity and Performance in Organizational Teams. *Administrative Science Quarterly*. 1992;37:634–665.
101. Manz CC, Sims HP. Leading Workers to Lead Themselves: The External Leadership of Self-Managing Work Teams. *Administrative Science Quarterly*. 1987;32:106–128.
102. Cox JF. *The effects of superleadership training on leader behavior, subordinate self-leadership behavior and subordinate self-leadership behavior and subordinate citizenship*. College Park: University of Maryland; 1994.
103. Pollack MM, Koch MA. Association of outcomes with organizational characteristics of neonatal intensive care units. *Critical Care Medicine*. 2003;31:1620–1629.
104. Dechairo-Marino AE, Jordan-Marsh M, Traiger G, et al. Nurse/physician collaboration: action research and the lessons learned. *J Nurs Adm*. 2001;31:223–232.
105. Gittell JH, Seidner R, Wimbush J. A Relational Model of How High-Performance Work Systems Work. *Organization Science*. 2010;21: 490–506.
106. Noel PH, Lanham HJ, Palmer RF, et al. The Importance of Relational Coordination and Reciprocal Learning for Chronic Illness Care within Primary Care Teams. *Health Care Management Review*. 2012.
107. Blegen MA, Sehgal NL, et al. Improving safety culture on adult medical units through multidisciplinary teamwork and communication interventions: the TOPS Project. *Quality & Safety in Health Care*. 2010;19:346–350.
108. Anderson A. Nurse-physician interaction and job satisfaction. *Nursing Management*. 1996;27:33–34.
109. Profit J, Etchegaray J, et al. The Safety Attitudes Questionnaire as a tool for benchmarking safety culture in the NICU. *Archives of Disease in Childhood-Fetal and Neonatal Edition*. 2012;97:F127–F132.
110. Kaya S, Barsbay S, et al. The Turkish version of the safety attitudes questionnaire: psychometric properties and baseline data. *Quality & Safety in Health Care*. 2010;19:572–577.
111. Rosenfeld PL. The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social Science & Medicine*. 1992;35:1343–1357.
112. Kalisch BJ, Lee KH. Nurse Staffing Levels and Teamwork: A Cross-Sectional Study of Patient Care Units in Acute Care Hospitals. *Journal of Nursing Scholarship*. 2011;43:82–88.
113. Kalisch BJ, Lee H, et al. Nursing staff teamwork and job satisfaction. *Journal of Nursing Management*. 2010a;18:938–947.
114. Kalisch BJ, Lee KH. The impact of teamwork on missed nursing care. *Nursing Outlook*. 2010b;58:233–241.
115. Sorra JS, Famolaro TND. Hospital Survey on Patient Safety Culture: User Comparative Database Report. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
116. Agency for Healthcare Quality Research. *Surveys on Patient Safety Culture*. Rockville, MA: Agency for Healthcare Research and Quality; 2012.
117. Klein KJ, Conn AB, Smith DB, et al. Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *J Appl Psychol*. 2001;86:3–16.