HARVARD | BUSINESS | SCHOOL

Measuring Teamwork in Health Care Settings: A Review of Survey Instruments

Melissa A. Valentine
Ingrid M. Nembhard
Amy C. Edmondson

Working Paper

11-116

April 12, 2012

Measuring Teamwork in Health Care Settings: A Review of Survey Instruments

Melissa A. Valentine
Harvard Business School
Boston, MA 02163
Tel: 617.852.8644
mvalentine@hbs.edu

Ingrid M. Nembhard
Yale University School of Medicine and School of Management
New Haven, CT 06520-8034
Tel: 203.785.3778
ingrid.nembhard@yale.edu

Amy C. Edmondson
Harvard Business School
Boston, MA 02163
Tel: 617.495.6732
aedmondson@hbs.edu

April 12, 2012

Measuring Teamwork in Health Care Settings: A Review of Survey Instruments

Abstract

Objective. To identify and review survey instruments used to assess dimensions of teamwork, a vital input to delivering quality care, so as to facilitate high quality research on this topic.

Data sources. The ISI Web of Knowledge database, which includes articles from MEDLINE, Social Science Citation Index, and Science Citation Index.

Study design. We conducted a systematic review of articles published before January 2010 to identify survey instruments used to measure teamwork and to assess their conceptual content, psychometric validity, and relationships to outcomes of interest.

Data extraction. We identified relevant articles using the search terms *team*, *teamwork*, *work groups*, or *collaboration*, in combination with *survey* or *questionnaire*.

Principal findings. We found 35 surveys that measured teamwork. Surveys differed in the dimensions of teamwork that they assessed. The most commonly assessed dimensions were communication, coordination and respect. Of the 35 surveys, nine met all of the criteria for psychometric validity and 13 have shown significant relationship to non-self-report outcomes.

Conclusions. "Teamwork" can refer to many different behavioral processes and emergent states, making it challenging and critical for researchers to develop a theory of teamwork consistent with their research context before selecting a survey. Psychometric validity is also vitally important. This review can help researchers identify high-quality teamwork surveys.

Key words. Teams, teamwork, psychometric properties, survey instruments

The use of teams has grown significantly in health care organizations, becoming a critical part of the way in which care is delivered (IOM 2001; JCAHO 1998). Today about 60% of U.S. primary care practices use team-based models (Schoen et al. 2009). The percentage reaches almost 100% in many other countries. According to many experts, team*work* is now an essential part of effective health care delivery, regardless of whether health professionals are assigned to designated teams, because of the increasing complexity of health care delivery (IOM 2003; Lemieux-Charles and McGuire 2006; Schmitt 2001). To deliver quality care, often a number of professionals with different expertise must work together.

Research suggests that the benefits of effective teamwork can be substantial. Recent studies show that higher team functioning is associated with better patient outcomes (Bower et al. 2003; Davenport et al. 2007; Shortell et al. 1991) and cost savings (Grumbach and Bodenheimer 2004). Scholars have theorized that these benefits accrue because better functioning teams make better quality decisions, cope better with complex tasks, produce more integrated care plans based on combined expertise, and better coordinate their actions (Dean, LaVallee, and McLaughlin 1999; Grumbach and Bodenheimer 2004; Wagner 2000).

Despite growing awareness of its potential benefits, effective teamwork is often lacking in health care organizations, with negative consequences for patients (IOM 2001). In a review of 54 malpractice incidents in an emergency department, 8 out of 12 deaths and 5 out of 8 permanent impairments were judged to be preventable if appropriate teamwork had occurred (Risser et al. 1999). The prevalence of teamwork failures has been attributed to several factors. First, a professional hierarchy exists in medicine, resulting in power and status differences within many care teams. When such differences exist, teamwork falters because both high and low status individuals fail to engage in open conversation for fear of negative consequences (e.g.,

embarrassment, disrupting the hierarchy) (Edmondson 1996; Lichtenstein et al. 2004; Nembhard and Edmondson 2006). Second, frequent transitions between caregivers due to shift-changes, patient transfers, or academic teaching schedules make coordination and teamwork more complicated (Wageman, Hackman, and Lehman 2005). Finally, teamwork requires dealing with the challenges of human relationships and different personalities, which can create process losses that overtake the benefits of working together (Steiner 1972).

These previous studies make it clear that teamwork may not happen naturally in health care, but it is critical for supporting quality care, quality improvement, patient safety, worker satisfaction, and cost-savings efforts. Supporting teamwork requires a strong theoretical and empirical understanding of what teamwork is, which depends in part on the appropriate measurement of teamwork. However, there has not yet been a systematic review of the survey instruments available for assessing teamwork in health care settings.

In this paper, we report the results of our systematic review of surveys examining teamwork. We focus on surveys as opposed to other methodologies for assessing teamwork (e.g., direct observation) because – despite being subject to well-known biases (e.g., response bias; see Paulhus (1991) for a discussion) – surveys are relatively easy to administer, are not resource intensive for large samples, and provide data that can be used to examine relationships between variables statistically. Our aim is to assist with survey selection by providing a comprehensive review of the dimensions of teamwork assessed by each survey as well as the psychometric validity of each survey. To facilitate understanding of the dimensions of teamwork that we ultimately assess, we begin by reviewing the concept of teamwork.

CONCEPTUAL BACKGROUND: WHAT IS TEAMWORK?

Even among highly-cited reviews of teamwork and team processes, there is no one

unifying theory of exact dimensions of teamwork (see for example Dickinson and McIntyre 1997; Ilgen et al. 2005; Marks, Mathieu, and Zaccaro 2001). Instead, the term "teamwork" encapsulates a broad set of behavioral processes that people use to accomplish interdependent work, as well as affective, cognitive and motivation states that emerge during the course of that work (Ilgen et al. 2005). Behavioral processes include actions such as communication, coordination, use of others' expertise, and helping. Emergent states include, for example, mutual respect and psychological safety. Behavioral processes and emergent states are distinct from permanent traits, group structures, or individual characteristics, and also from task work (e.g., interactions with tools and systems) (Bowers, Braun, and Morgan 1997).

Because the term "teamwork" is used as a catchall to refer to a number of behavioral processes and emergent states, measures of teamwork can be expected to be diverse. Some of this diversity represents an opportunity for more cumulative research, but some of the diversity reflects substantive differences on important factors. Those factors include the purpose of the research, the type of team being studied, and the type of task being studied.

Research Purpose

First, the dimensions of teamwork assessed will depend on the purpose of the research. For example, the purpose of the research might be to develop and test theory about specific behavioral processes. In this case, a more narrow and precise conceptualization of certain aspects of a team's behavioral processes would be adopted (e.g., Edmondson 1999). Alternatively, the aim of the research might be to develop a broad understanding of collaborative work, including all of the behaviors and emergent states that might matter when people work interdependently, in which case a more broad and comprehensive collection of behaviors would be assessed (e.g., Hoegl and Gemuenden 2001). Many studies assess behavioral processes and

emergent states as part of developing a full model of team effectiveness that includes measures of organizational context, team design, team composition, team structure, and task design, as well as measures of behavioral processes and emergent states (e.g., Campion, Medsker, and Higgs 1993; Pinto, Pinto, and Prescott 1993; Wageman et al. 2005).

Team Type

Second, the dimensions of teamwork assessed by a survey might vary according to the type of team being studied (Hackman and Katz 2010; Hollenbeck, Beersma, and Schouten 2012). Some teams have stable, clearly delimited membership. In such teams, measuring behaviors through which interdependent tasks are accomplished, like monitoring progress or formulating strategy, might be most appropriate (e.g., Wageman et al. 2005). However, such behaviors may not be relevant in situations where people are not organized into a formal team but must engage in effective teamwork with shifting partners. For example, nurses and physicians in an intensive care unit work interdependently to care for patients, typically without being in formal teams. In such settings, assessing behaviors like cooperation and communication in the broader unit might be more theoretically and empirically relevant than assessing behaviors like monitoring and strategy formulation, which are more meaningful within formal teams (e.g., Shortell et al. 1991).

Task Type

Finally, the dimensions of teamwork assessed in a survey might vary according to the type of task being studied (Stewart 2006; Stewart and Barrick 2000). Some tasks are more conceptual and require planning, strategizing, or diagnosing; some are more behavioral, requiring physical actions. For conceptual tasks, teamwork might mean effectively drawing on and combining various people’s expertise. For behavioral tasks, teamwork might mean coordinating timing and helping each other.

Each of the above factors – research purpose, team type, and task type – can influence the specific dimensions of teamwork measured in a survey. There are multiple dimensions of teamwork, giving rise to a variety of surveys. Researchers must consider which factors are most salient to their research question, and then organize their conceptualizations of teamwork and its correlates into a nomological network that captures causal logic (Cronbach and Meehl 1955; Dickinson and McIntyre 1997). This conceptual work should guide the selection of survey measures and instruments.

In the next section, we discuss the methods we used to assess the conceptual content and psychometric validity of existing teamwork surveys, with the aim of assisting researchers and practitioners interested in teamwork with the selection of an appropriate survey for their work.

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[i] using the keywords: *team*, *teamwork*, *work groups*, or *collaboration*, in combination with *survey* or *questionnaire.* In addition to ISI, we searched the references of five highly-cited literature reviews on teams (Bettenhausen 1991; Cohen and Bailey 1997; Guzzo and Dickson 1996; Holland, Gaston, and Gomes 2000; Lemieux-Charles and McGuire 2006). We examined every referenced article to determine whether the authors used surveys to measure teamwork. We then examined the references from all of the articles identified using the above two strategies (ISI search and review articles) to find any additional articles that used surveys to measure teamwork.

In total, we examined over 1,800 articles in management, social science, medicine, and health services research journals. We excluded the vast majority of these articles from further

review because they were not published in peer-reviewed journals, did not empirically assess teamwork, or reported on studies that used methods other than surveys to assess teamwork, such as interviews (Makowsky et al. 2009; Slonski-Fowler and Truscott 2004), direct observation (Healey et al. 2008), video analysis (Mackenzie and Xiao 2003) or behavioral marker systems (Malec et al. 2007; Mathieu et al. 2000). We also excluded surveys that used an individual level of analysis (e.g., Weiss and Davis 1985), that measured development over time (e.g., Wheelan and Hochberger 1996), or did not measure behavior (e.g., Gibson 2003).

We retained 35 articles in our sample 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 two ways. First, we reviewed the dimensions of teamwork assessed by the surveys. We then assessed the psychometric strength of each survey and also whether the survey had an established relationship with a non-self-report outcome.

Reviewing the Dimensions of Teamwork Assessed

Because the dimensions assessed by each survey likely relates to the developers' research purpose and the type of team or task studied, we divided surveys by research purpose and team type, and then qualitatively assessed the dimensions of teamwork contained in each survey. We first distinguished between surveys developed for the purpose of creating models of team effectiveness versus those developed for other purposes. All of the surveys developed to test models of team effectiveness were developed for bounded teams.

We next divided the surveys developed for other purposes by the type of team described (i.e., bounded teams versus larger, unbounded workgroups like units or departments). For each group of surveys, the first two authors and a research assistant independently reviewed each item in every survey and categorized each as a behavioral process, an emergent state, or other. We

then further categorized each using the sub-categories of behavioral processes and emergent states that emerged during our review. Our intent in reviewing (and presenting) the dimensions of teamwork assessed by each survey was to help researchers identify a survey relevant to the theory of teamwork developed for their new study.

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 four criteria. Although these criteria are well-established and generally accepted, we note that what is ultimately acceptable depends on research setting and purpose (Lance, Butts, and Michels 2006; LeBreton and Senter 2008). However, at a minimum, a good survey will perform well with respect to all four criteria.

1. *Internal consistency or reliability*. Internal consistency refers to the correlation between items in a survey measure. In a good survey, the correlation between measure items is high, suggesting that items within the measure capture the same latent construct. A commonly used statistic for assessing internal consistency is Cronbach's alpha, which ranges between negative infinity and 1 (Cronbach 1951). In applied settings where decisions are to be made based on scores, experts note that a value of 0.9 is "the minimum that should be tolerated" (Nunnally 1976 pg. 245). However, for early stage research and newly developed surveys, a minimum value of 0.7 is generally considered acceptable. It indicates moderate consistency between items (70% of variance is true score variance, 30% is random measurement error variance) (Lance et al. 2006; Nunnally 1976).
2. *Interrater agreement and reliability*. A good survey will elicit similar responses about the phenomenon of interest (e.g., teamwork) from different judges (e.g.,

each person in the team). Both interrater agreement (IRA) and interrater reliability (IRR) assess the level of similarity between responses provided by different judges. However, they differ in how they define similarity. IRA focuses on the absolute consensus between judges, while IRR focuses on relative consistency between judges (Bliese 2000; LeBreton and Senter 2008) 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 (James et al. 1984). A r_{wg} value of 0.7 is often cited as the minimum acceptable value, although this has been debated (Lance et al. 2006).

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 a metric for both criteria (LeBreton and Senter 2008). ICC values greater than zero indicate similarity (Shrout and Fleiss 1979). Some have argued that due to the difference in focus, both IRA and IRR should be reported as standard practice (Klein et al. 2000).

Note that IRA and IRR are particularly important for surveys measuring phenomena such as teamwork that are believed to 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 (e.g., ICC) metrics

should be used to determine whether aggregation is warranted. Results of within and between analysis (WABA), which uses an analysis of variance (ANOVA) to test whether variation between groups is greater than variation within a group, can also be used to justify aggregation.

3. *Discriminant validity.* Discriminant validity refers to the extent to which items or measures within the survey that are theoretically dissimilar diverge. When a survey measure captures a distinct construct, the results of exploratory and confirmatory factor analysis will show that all items in the scale belong to one "factor" and have limited association with other factors/constructs. Additionally, items expected to be unrelated to the focal construct will not correlate with it. To provide evidence of discriminant validity based on factor analysis, several results should be reported: the number of distinct factors, the percentage of variance explained by the factor structure, the values of factor loadings (ideally, greater than 0.40), or eigenvalues (ideally, greater than 1.0). Ideally, many theoretical constructs will also be tested against each other during measure development. See Cronbach and Meehl (1955) for an in-depth discussion of this process.
4. *Content (or external) validity.* The content validity criterion requires that a survey be demonstrated to actually 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" (Denzin 1978). Researchers triangulate by comparing survey results to data obtained via observation, semi-structured interviews, qualitative work, and/or expert or respondent review of the survey (Edmondson and

McManus 2007; Jick 1979). This comparison minimizes the risk that a survey captures a priori assumptions about what is important in the construct, rather than the true dimensions of the construct.

We also report the number of items in a survey measure, and the context in which the survey was developed. This information about the context in which the survey was originally developed may indicate how much the survey will need to be adapted for use in a new setting.

After evaluating the psychometric strength of each survey, we lastly examined the peer-reviewed literature related to each survey to determine whether existing research had documented a relationship between each survey measure and a non-self-reported outcome (e.g., clinical outcome or manager-rater team effectiveness).

RESULTS

Each of the 35 peer-reviewed articles reported the development or use of a survey measuring teamwork. The surveys, all of which were published during the last 20 years (1991-2011), were less likely to appear in health services or medical journals (16 surveys) than in general management journals (19 surveys). Only one, the relational coordination survey, was published in both a health services and general management journal (Gittell 2002; Gittell et al. 2000).

The Dimensions of Teamwork in Surveys

Of the 35 surveys developed to measure teamwork, nine were developed as part of a team effectiveness model. Thus, other core elements of the proposed model – organizational context, team design, task design, and team performance – were assessed along with teamwork (Table 1).

Of the remaining 26 surveys, 12 were used to assess teamwork in bounded teams, and 14 were used to assess teamwork in larger, unbounded workgroups like units or departments.

Across the 12 surveys 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). Two of the 12 surveys did not assess emergent states. The surveys that assessed the most dimensions were Hoegl (2001) and Anderson (1998).

Of the 14 surveys that examined teamwork in larger, unbounded groups, 12 focused on behavioral processes and emergent states and two (Heinemann et al. 1999; Hojat et al. 1999) focused on attitudes towards teamwork. The 12 surveys that assessed behaviors and emergent states in 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. The surveys that assessed the most dimensions of teamwork within unbounded workgroups were Adams (1995) and Kalisch (2010); both assessed 11 dimensions of teamwork.

On average, surveys developed for larger, unbounded work groups assessed more dimensions of teamwork. These surveys also did not assess group cohesion/shared identity, which was commonly assessed in bounded teams. Across both team types, there was more focus on behavioral processes than emergent states and communication and coordination were the most commonly assessed behavioral processes.

The Psychometric Validity of Teamwork Surveys

Only 14 of the 35 teamwork surveys (40%) were reported with the full set of psychometric properties that we evaluated and of those, nine satisfied the minimum standards for all of these criteria (Table 4). Those that completely satisfied the minimum standards are indicated by an "X" in a shaded square in the first column of Tables 1-4. 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 (e.g., Shortell (1991) reported an alpha value of 0.64, which is just below the threshold of 0.70).

Of the 24 that did not report values for all of the psychometric properties that we evaluated, 22 did not report interrater agreement or reliability, one (Gittell 2002) did not report discriminant validity, and one (Brannick, Roach, and Salas 1993) did not clearly report either interrater agreement or discriminant validity.

The Relationship between Surveys and Outcomes of Interest

Of the 35 teamwork surveys identified, 13 had documented relationships with non-self-reported outcomes. Five with clinical outcomes (Alexander et al. 2005; Baggs 1994; Gittell et al. 2000; Sexton et al. 2006; Sorra 2004), six with a non-clinical performance metric (Campion et al. 1993; Edmondson 1999; Hoegl and Gemuenden 2001; Kalisch et al. 2010; Vinokur-Kaplan 1995; Wageman 2005), and two with both clinical and non-clinical outcomes (Anderson and West 1998; Shortell et al. 1991). Of the remaining 22 surveys, nine had not been examined relative to an outcome (i.e., 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 (e.g., organizational culture (Strasser 2002)).

Notably, the 13 surveys with a documented relationship to a non-self-reported outcome were more likely to be reported with the full set of psychometric properties: eight of these surveys (60%) were reported with the full set of psychometric properties we evaluated, and 4 of these satisfied the minimum standard for the four criteria that we assessed (see 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 care delivery, and patient safety. Many surveys have been developed to assess teamwork and there is considerable variation in the dimensions of teamwork measured across surveys. Some of this reflects different research focus (i.e., developing a model of team effectiveness versus testing specific antecedents of teamwork) or team type (i.e., bounded or unbounded). However, some dimensions of teamwork appeared consistently, even across the different foci and team types: communication, coordination, use of all members' expertise, and respect, which suggests that these may be core dimensions.

There is also variation in the quality of teamwork measures. Only nine of the 35 surveys satisfied standard psychometric criteria, and only four of those have been significantly associated with non-self-reported outcomes. Several other surveys missed the cut-offs values by relatively narrow margins. The majority of the surveys fail to either meet or report the standard psychometric criteria expected of survey instruments. Evidence for two of the four criteria — interrater agreement and content validity — were rarely reported. Both, along with internal consistency and discriminant validity, are critical to establishing the statistical validity and reliability of surveys. Interrater agreement demonstrates how well a measure gathers reliable information, and discriminant/content validity is important for assessing whether it captures substantive reality (Jick 1979). The absence of this information makes it difficult for others to evaluate the appropriateness of surveys or measures for their use. At least one indicator of each of the four 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.

It is noteworthy that of the 35 teamwork surveys identified, only 13 had a demonstrated

relationship to non-self-reported outcomes. For many of the remaining surveys, this was because they did not include objective outcomes in the study rather than because of a finding of no association. Further, our findings suggest a general propensity for researchers to develop new surveys for projects, rather than adopt or adapt existing surveys, potentially limiting cumulative knowledge. Many of the hypothesized effects of teamwork thus remain underexplored. With respect to failing to use existing surveys, it may be that they were inappropriate for the research setting or that robust measures to test other aspects of teamwork have not been available. However, as research on teamwork advances, the field would be well served by studies utilizing existing, psychometrically valid surveys to promote cumulative knowledge of teamwork.

Limitations

A common limitation in review articles comes from having to define a search area, and in so doing losing other valuable and relevant information. Our review focused on surveys that assess teamwork, but we note that for nearly every specific dimension of teamwork assessed in these articles there is also a rich and varied research literature specific to that dimension (e.g., communication, decision-making, conflict management). We did not include these dimensions as search terms and did not include specific measures of specific dimensions of teamwork in our review for practical reasons (e.g., space constraints). But researchers may find value in further searching the team literature for specific survey measures if one behavioral process seems 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. Klein et al. (2001) showed that the use of a group rather than individual referent increased within-group agreement in response to descriptive items but decreased within-group agreement

in response to evaluative items. Thus, wording-induced bias is important to consider, particularly in assessing IRR. We did not assess surveys for this potential source of bias because a widely-agreed upon assessment does not yet exist. Also, we were not able to assess whether surveys tested for the discriminant validity of all their measures against appropriate counter measures. Establishing discriminant validity often requires more testing of theoretically similar constructs than is typically reported in articles.

IN SUMMARY: CHOOSING A SURVEY INSTRUMENT

This article is intended to help researchers or practitioners who ask: which is the best teamwork survey to use in future work? The answer will depend on a number of factors.

First and foremost, there should be conceptual consistency between the survey selected and the theory of action for the research context (Cronbach and Meehl 1955). In other words, survey selection begins with an understanding of the teamwork dimensions applicable to the specific context and then a review of the instruments that measure those dimensions (Dickinson and McIntyre 1997). The theory will depend on several things including whether teamwork is being enacted in a bounded team or larger group and on the nature of the task. (For researchers who seek more background for developing a theory of teamwork, articles by Hackman (1987), Cohen and Bailey (1997), Ilgen (2005), Salas (2008) and Kozlowski (2008) are helpful, as is a review of team effectiveness specific to health care settings by LeMieux-Charles (2006)). Note that for the conceptual background cited above to be applicable and relevant, it is important for the workers being assessed to be interdependent (see for example Sprigg, Jackson, and Parker 2000), which is not always the case in health care settings.

Second, researchers may need to consider whether and how to adapt an existing survey to a new setting. The theory of teamwork may look different in an ICU 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 easier it will be to use that survey is diverse settings. However, it might be more difficult to assess the particular processes in the causal pathway between teamwork and team performance if the teamwork survey is too general.

Third, the survey should satisfy established criteria for psychometric validity. Using psychometrically valid surveys enables the user to have greater confidence in results. Lastly, users should consider administrative constraints. Surveys vary considerably in the number of items they contain (range: 6-82) and longer surveys may limit the possibility of assessing additional constructs in the same survey.

This paper aims to assist the selection process by reviewing the dimensions of teamwork and psychometric properties of existing teamwork surveys. We hope that it helps scholars to identify high quality existing 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, found in existing surveys, when possible to facilitate the development of cumulative knowledge about teamwork. Though efforts were made to identify as many existing teamwork surveys as possible, we cannot claim to have been exhaustive. However, we believe that the criteria set forth in this article should be considered standard research practice, and as such the surveys that we identified are worthy of attention.

Surveys: (Adams et al. 1995; Alexander et al. 2005; Anderson and West 1998; Baggs 1994; Bateman 2002; Brannick et al. 1993; Campion et al. 1993; Copnell et al. 2004; Denison, Hart, and Kahn 1996; Doolen, Hacker, and Van Aken 2003; Edmondson 1999; Friesen et al. 2008; Gittell 2002; Hauptman and Hirji 1999; Heinemann et al. 1999; Hoegl and Gemuenden 2001; Hojat et al. 1999; Hutchinson et al. 2006; Kahn and McDonough 1997; Kalisch et al. 2010; La Duckers, Wagner, and Groenewegen 2008; Masse et al. 2008; Millward and Jeffries 2001; Pearce and Sims 2002; Pinto et al. 1993; Seers 1989; Senior and Swailes 2007; Sexton et al. 2006; Shortell et al. 1991; Sorra 2004; Strasser et al. 2002; Ushiro 2009; van Beuzekom, Akerboom, and Boer 2007; Vinokur-Kaplan 1995; Wageman et al. 2005)

References

Adams, A., S. Bond, and S. Arber. 1995. “Development and validation of scales to measure organisational features of acute hospital wards.” *International Journal of Nursing Studies* 32(6): 612-27.

Alexander, J. A., R. Lichtenstein, K. Jinnett, R. Wells, J. Zazzali, and D. W. Liu. 2005. “Cross-functional team processes and patient functional improvement.” *Health Services Research* 40(5): 1335-55.

Anderson, N. R. and M. A. West. 1998. “Measuring climate for work group innovation: development and validation of the team climate inventory.” *Journal of Organizational Behavior* 19(3): 235-58.

Baggs, J. G. 1994. “Development of an Instrument to Measure Collaboration and Satisfaction about Care Decisions ” *Journal of Advanced Nursing* 20(1): 176-82.

Bateman, B., Wilson, F., Bingham, D. 2002. “Team effectiveness: development of an audit questionnaire.” *The Journal of Management Development* 21(3/4): 215-26.

Bettenhausen, K. L. 1991. “Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed.” *Journal of Management* 17(2): 345.

Bliese, P. D. 2000. “Within-Group Agreement, Non-Independence, and Reliability: Implications for Data Aggregation and Analysis.” In *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions and New Directions*, edited by K. J. Klein and S. W. J. Kozlowski, pp. 349-81. San Francisco: Jossey-Bass.

Bower, P., S. Campbell, C. Bojke, and B. Sibbald. 2003. “Team structure, team climate and the quality of care in primary care: an observational study.” *Quality & Safety in Health Care* 12(4): 273-79.

Bowers, C. A., C. C. Braun, and B. B. Morgan. 1997. “Team workload: Its meaning and measurement.” In *Team performance and measurement: Theory, methods, and applications*, edited by M. T. Brannick, E. Salas, and C. Prince, pp. 85-108. Mahwah, NJ: Lawrence Erlbaum Associates.

Brannick, M. T., R. M. Roach, and E. Salas. 1993. “Understanding Team Performance: A Multimethod Study.” *Human Performance* 6(4): 287.

Campion, M. A., G. J. Medsker, and A. C. Higgs. 1993. “Relations between work group characteristics and effectiveness - implications for designing effective work groups ” *Personnel Psychology* 46(4): 823-50.

Cohen, S. G. and D. E. Bailey. 1997. “What makes teams work: Group effectiveness research from the shop floor to the executive suite.” *Journal of Management* 23(3): 239-90.

Copnell, B., L. Johnston, D. Harrison, A. Wilson, A. Robson, C. Mulcahy, L. Ramudu, G. McDonnell, and C. Best. 2004. “Doctors' and nurses' perceptions of interdisciplinary collaboration in the NICU, and the impact of a neonatal nurse practitioner model of practice.” *Journal of Clinical Nursing* 13(1): 105-13.

Cronbach, L. J. and P. E. Meehl. 1955. “Construct Validity in Psychological Tests ” *Psychological Bulletin* 52(4): 281-302.

Davenport, D. L., W. G. Henderson, C. L. Mosca, S. F. Khuri, and R. M. Mentzer. 2007. “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.” *Journal of the American College of Surgeons* 205(6): 778-84.

Dean, P. J., R. LaVallee, and C. P. McLaughlin. 1999. “Teams at the Core of Continuous Learning.” In *Continuous Quality Improvement in Health Care: Theory Implementation and Applications*, edited by C. P. McLaughlin. Gaithersburg: Aspen Publishers.

Denison, D. R., S. L. Hart, and J. A. Kahn. 1996. “From chimneys to cross-functional teams: Developing and validating a diagnostic model.” *Academy of Management Journal* 39(4): 1005-23.

Denzin, N. K. 1978. *The Research Act*. New York: McGraw-Hill.

Dickinson, T. and R. McIntyre. 1997. “A conceptual framework for teamwork measurement.” In *Team Performance Assessment and Measurement: Theory, Methods, and Applications*, edited by M. T. Brannick, E. Salas, and C. Prince, pp. 19-43. Mahwah, NJ: Lawrence Erlbaum.

Doolen, T. L., M. E. Hacker, and E. M. Van Aken. 2003. "The impact of organizational context on work team effectiveness: A study of production team." *Ieee Transactions on Engineering Management* 50(3): 285-96.

Edmondson, A. 1996. "Learning from Mistakes is Easier Said Than Done: Group and Organizational Influences on the Detection and Correction of Human Error." *Journal of Applied Behavioral Science* 32(1): 5-28.

Edmondson, A. 1999. "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly* 44(2): 350-83.

Edmondson, A. C. and S. E. McManus. 2007. "Methodological fit in management field research." *Academy of Management Review* 32: 1155-79.

Friesen, L. D., A. Vidyarthi, R. B. Baron, and P. P. Katz. 2008. "Factors Associated with Intern Fatigue." *Journal of General Internal Medicine* 23(12): 1981-86.

Gibson, C., Zellmer-bruh, M., Schwab D. 2003. "Team Effectiveness in Multinational Organizations: Evaluation Across Contexts." *Group and Organization Management* 28(4): 444-74.

Gittell, J. H. 2002. "Coordinating mechanisms in care provider groups: Relational coordination as a mediator and input uncertainty as a moderator of performance effects." *Management Science* 48(11): 1408-26.

Gittell, J. H., K. M. Fairfield, B. Bierbaum, W. Head, R. Jackson, M. Kelly, R. Laskin, S. Lipson, J. Siliski, T. Thornhill, and J. Zuckerman. 2000. "Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay - A nine-hospital study of surgical patients." *Medical Care* 38(8): 807-19.

Grumbach, K. and T. Bodenheimer. 2004. "Can health care teams improve primary care practice?" *Jama-Journal of the American Medical Association* 291(10): 1246-51.

Guzzo, R. A. and M. W. Dickson. 1996. "Teams in Organizations: Recent Research on Performance and Effectiveness." *Annual Review of Psychology* 47.

Hackman, J. R. 1987. "The design of work teams." In *Handbook of organizational behavior*, edited by J. Lorsch. Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R. and N. Katz. 2010. "Group behavior and performance." In *Handbook of Social Psychology*, edited by S. Fiske. New York: Wiley.

Hauptman, O. and K. K. Hirji. 1999. "Managing integration and coordination in cross-functional teams: an international study of Concurrent Engineering product development." *R & D Management* 29(2): 179-91.

Healey, A. N., S. Olsen, R. Davis, and C. A. Vincent. 2008. "A method for measuring work interference in surgical teams." *Cognition, Technology & Work* 10(4): 305-12.

Heinemann, G. D., M. H. Schmitt, M. P. Farrell, and S. A. Brallier. 1999. "Development of an attitudes toward health care teams scale." *Evaluation & the Health Professions* 22(1): 123-42.

Hoegl, M. and H. G. Gemuenden. 2001. "Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence." *Organization Science* 12(4): 435-49.

Hojat, M., S. K. Fields, J. J. Veloski, M. Griffiths, M. J. M. Cohen, and J. D. Plumb. 1999. "Psychometric properties of an attitude scale measuring physician-nurse collaboration." *Evaluation & the Health Professions* 22(2): 208-20.

Holland, S., K. Gaston, and J. Gomes. 2000. "Critical success factors for cross-functional teamwork in new product development." *International Journal of Management Reviews* 2(3): 231-59.

Hollenbeck, J. R., B. Beersma, and M. E. Schouten. 2012. "Beyond team types and taxonomies: A dimension scaling conceptualization for team description " *Academy of Management Review* 37(1): 82-106.

Hutchinson, A., K. L. Cooper, J. E. Dean, A. McIntosh, M. Patterson, C. B. Stride, B. E. Laurence, and C. M. Smith. 2006. "Use of a safety climate questionnaire in UK health care: factor structure, reliability and usability." *Quality & Safety in Health Care* 15(5): 347-53.

Ilgen, D. R., J. R. Hollenbeck, M. Johnson, and D. Jundt. 2005. "Teams in Organizations: From Input-Process-Output Models to IMOI Models." *Annual Review of Psychology* 56(1): 517-43.
IOM. 2001. "Crossing the Quality Chasm: A New Health System for the 21st Century." *Institue of Medicine*. Washington DC: New Academy Press.
IOM. 2003. "Keeping Patients Safe: Transforming the Work Environment of Nurses." Washington DC: Institute of Medicine.
JCAHO. 1998. "Sentinel events: evaluating cause and planning improvement." Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.
Jick, T. D. 1979. "Mixing Qualitative and Quantitative Methods - Triangulation in Action " *Administrative Science Quarterly* 24(4): 602-11.
Kahn, K. B. and E. F. McDonough. 1997. "An empirical study of the relationships among co-location, integration, performance, and satisfaction." *Journal of Product Innovation Management* 14(3): 161-78.
Kalisch, B. J., H. Lee, and E. Salas. 2010. "The Development and Testing of the Nursing Teamwork Survey." *Nursing Research* 59(1): 42-50.
Kozlowski, S. 2008. "Team processes and team effectiveness: Fifty years of progress and prospects for the future." *International Journal of Psychology* 43(3-4): 366-67.
La Duckers, M., C. Wagner, and P. P. Groenewegen. 2008. "Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives." *Bmc Health Services Research* 8: 9.
Lance, C. E., M. M. Butts, and L. C. Michels. 2006. "The sources of four commonly reported cutoff criteria - What did they really say?" *Organizational Research Methods* 9(2): 202-20.
LeBreton, J. M. and J. L. Senter. 2008. "Answers to 20 questions about interrater reliability and interrater agreement." *Organizational Research Methods* 11(4): 815-52.
Lemieux-Charles, L. and W. L. McGuire. 2006. "What do we know about health care team effectiveness? A review of the literature." *Medical Care Research and Review* 63(3): 263-300.
Lichtenstein, R., J. A. Alexander, J. F. McCarthy, and R. Wells. 2004. "Status differences in cross-functional teams: Effects on individual member participation, job satisfaction, and intent to quit." *Journal of Health and Social Behavior* 45(3): 322-35.
Mackenzie, C. F. and Y. Xiao. 2003. "Video techniques and data compared with observation in emergency trauma care." *Quality & Safety in Health Care* 12: II51-II57.
Makowsky, M. J., T. J. Schindel, M. Rosenthal, K. Campbell, R. T. Tsuyuki, and H. M. Madill. 2009. "Collaboration between pharmacists, physicians and nurse practitioners: A qualitative investigation of working relationships in the inpatient medical setting." *Journal of Interprofessional Care* 23(2): 169-84.
Malec, J. F., L. C. Torsher, W. F. Dunn, D. A. Wiegmann, J. J. Arnold, D. A. Brown, and V. Phatak. 2007. "The Mayo High Performance Teamwork Scale: Reliability and Validity for Evaluating Key Crew Resource Management Skills." *Simulation in Healthcare* 2(1): 4-10 10.1097/SIH.0b013e31802b68ee.
Marks, M. A., J. E. Mathieu, and S. J. Zaccaro. 2001. "A temporally based framework and taxonomy of team processes." *Academy of Management Review* 26(3): 356-76.
Masse, L. C., R. P. Moser, D. Stokols, B. K. Taylor, S. E. Marcus, G. D. Morgan, K. L. Hall, R. T. Croyle, and W. M. Trochim. 2008. "Measuring collaboration and transdisciplinary integration in team science." *American Journal of Preventive Medicine* 35(2): S151-S60.
Mathieu, J. E., T. S. Heffner, G. F. Goodwin, E. Salas, and J. A. Cannon-Bowers. 2000. "The influence of shared mental models on team process and performance." *Journal of Applied Psychology* 85(2): 273-83.
Millward, L. J. and N. Jeffries. 2001. "The team survey: a tool for health care team development." *Journal of Advanced Nursing* 35(2): 276-87.
Nembhard, I. M. and A. C. Edmondson. 2006. "Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams." *Journal of Organizational Behavior* 27(7): 941-66.
Nunnally, J. C. 1976. *Psychometric theory*. New York: McGraw-Hill.

Paulhus, D. 1991. “Measurement and Control of Response Bias.” In *Measures of personality and social pscyhological attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman. San Diego: Academic Press Inc.

Pearce, C. L. and H. P. Sims, Jr. 2002. “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, Research, and Practice* 6(2): 172-97.

Pinto, M. B., J. K. Pinto, and J. E. Prescott. 1993. “Antecedents and Consequences of Project Team Cross-functional Cooperation ” *Management Science* 39(10): 1281-97.

Risser, D. T., M. M. Rice, M. L. Salisbury, R. Simon, G. D. Jay, and S. D. Berns. 1999. “The potential for improved teamwork to reduce medical errors in the emergency department.” *Annals of Emergency Medicine* 34(3): 373-83.

Salas, E., N. J. Cooke, and M. A. Rosen. 2008. “On teams, teamwork, and team performance: Discoveries and developments.” *Human Factors* 50(3): 540-47.

Schmitt, M. H. 2001. “Collaboration improves the quality of care: methodological challenges and evidence from US health care research.” *Journal of Interprofessional Care* 15(1): 47-66.

Schoen, C., R. Osborn, M. M. Doty, D. Squires, J. Peugh, and S. Applebaum. 2009. “A Survey Of Primary Care Physicians In Eleven Countries, 2009: Perspectives On Care, Costs, And Experiences.” *Health Affairs* 28(6): W1171-W83.

Seers, A. 1989. “Team-Member Exchange Quality - A New Construct for Role-Making Research ” *Organizational Behavior and Human Decision Processes* 43(1): 118-35.

Senior, B. and S. Swailes. 2007. “Inside management teams: Developing a teamwork survey instrument.” *British Journal of Management* 18(2): 138-53.

Sexton, J. B., R. L. Helmreich, T. B. Neilands, K. Rowan, K. Vella, J. Boyden, P. R. Roberts, and E. J. Thomas. 2006. “The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research.” *Bmc Health Services Research* 6: 10.

Shortell, S. M., D. M. Rousseau, R. R. Gillies, K. J. Devers, and T. L. Simons. 1991. “Organizational Assessment in Intensive-Care Units (ICUs) - Construct Development, Reliability, and Validity of the ICU Nurse-Physician Questionnaire ” *Medical Care* 29(8): 709-26.

Slonski-Fowler, K. E. and S. D. Truscott. 2004. “General education teachers' perceptions of the prereferral intervention team process.” *Journal of Educational and Psychological Consultation* 15(1): 1-39.

Sorra, J., Nieva, V. 2004. “Hospital Survey on Patient Safety Culture.” *AHRQ Publication No. 04-0041*. Agency for Healthcare Research and Quality.

Sprigg, C. A., P. R. Jackson, and S. K. Parker. 2000. “Production teamworking: The importance of interdependence and autonomy for employee strain and satisfaction.” *Human Relations* 53(11): 1519-43.

Steiner, I. 1972. *Group Process and Productivity*: Academic Press Inc.

Stewart, G. L. 2006. “A meta-analytic review of relationships between team design features and team performance.” *Journal of Management* 32(1): 29-55.

Stewart, G. L. and M. R. Barrick. 2000. “Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type.” *Academy of Management Journal* 43(2): 135-48.

Strasser, D. C., S. J. Smits, J. A. Falconer, J. S. Herrin, and S. E. Bowen. 2002. “The influence of hospital culture on rehabilitation team functioning in VA hospitals.” *Journal of Rehabilitation Research and Development* 39(1): 115-25.

Ushiro, R. 2009. “Nurse-Physician Collaboration Scale: development and psychometric testing.” *Journal of Advanced Nursing* 65(7): 1497-508.

van Beuzekom, M., S. P. Akerboom, and F. Boer. 2007. “Assessing system failures in operating rooms and intensive care units.” *Quality & Safety in Health Care* 16(1): 45-50.

Vinokur-Kaplan, D. 1995. “Treatment Teams that work (and those that don't): An application of Hackman's group effectiveness model to interdisciplinary teams in psychiatric hospitals.” *Journal of Applied Behavioral Science* 31(3).

Wageman, R., J. R. Hackman, and E. Lehman. 2005. "Team Diagnostic Survey: Development of an Instrument." *The Journal of Applied Behavioral Science* 41(4): 373-98.
Wageman, R., Hackman, J.R., Lehman E. 2005. "Team Diagnostic Survey: Development of an Instrument." *Journal of Applied Behavioral Science* 41(4): 373-98.
Wagner, E. H. 2000. "The role of patient care teams in chronic disease management." *British Medical Journal* 320(7234): 569-72.
Weiss, S. J. and H. P. Davis. 1985. "Validity and Reliability of the Collaborative Practices Scale." *Nursing Research* 34(5): 299-305.
Wheelan, S. A. and J. M. Hochberger. 1996. "Validation studies of the group development questionnaire." *Small Group Research* 27(1): 143-70.

i The ISI Web of Knowledge search includes MEDLINE, the Social Science Citation Index, and the Science Citation Index. Two social science journals in the bibliographies of articles we identified were not in the ISI Web of Knowledge. We requested that they be added and the sponsors of ISI added them.

Table 1. Teamwork Dimensions Assessed in Full Models of Team Effectiveness[1]

	Psychometric Validity	Related to Outcomes	INPUTS			MEDIATORS (teamwork)		Outputs
			Organizational Context	Team Design	Team Task Design	Behavioral Processes	Emergent States	
Campion 1993		X				Workload sharing Communication	Social support Potency	
Denison 1996						Workload sharing Use of Expertise Strategy	Norms Teamwork Values	
Vinokur-Kaplan 1995		X				Effort Use of expertise Strategy		
Edmondson 1999	**X**	X				Team learning behaviors	Psychological safety Team efficacy	
Doolen 2003	**X**					Information sharing Team processes		
Wageman 2005		X				Effort Use of expertise Strategy Social interactions		
Senior 2007						Task interactions	Social support	
Pinto 1993						Cooperation		
Bateman 2002	**X**					Use of resources	Team synergy	

[1]Surveys listed in rows, sorted by number of dimensions assessed. Team effectiveness dimensions listed in columns, sorted by Input-Mediator-Output categories (Ilgen 2005). Specific dimensions listed in a full table available online. An X in the first column indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the second column indicates that a survey has an established relationship with a non-self-report outcome.

Table 2. Dimensions of Teamwork Assessed by Surveys Developed for Bounded Teams[2]

			Behavioral Processes									Emergent States					
												Affective				Cognitive	
	Psychometric Validity	**Related to Outcomes**	General teamwork quality	Communication	Coordination (mutual adjustment)	Collaboration	Use of all members' expertise	Help each other/share workload	Shared decision making	Active conflict management	Effort	Respect	Group cohesion/shared identity	Social support	Psychological safety	Role responsibility understanding	Shared objectives
Anderson 1998	X	X		■	■	■	■	■	■			■	■	■	■		■
Hoegl 2006	**X**	X		■	■	■	■	■	■	■	■	■	■	■			
Strasser 2002				■		■	■			■		■			■	■	■
Millward 2001	**X**			■	■		■	■				■				■	■
Alexander 2005	**X**	X		■					■			■	■		■		
Brannick 1993				■	■	■							■				
Seers 1995				■	■			■								■	
Hauptman 1999				■	■	■	■										
Kahn 1997			■			■											■
LaDuckers 2008				■	■											■	
Friesen 2008												■	■	■			
Pearce 2002	**X**		■	■													

[2] Surveys listed in rows, sorted by number of teamwork dimensions assessed. Teamwork dimensions listed in columns, sorted within categories by number of surveys by which each dimension was assessed. An X in the first column indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the second column indicates that a survey has an established relationship with a non-self-reported outcome. (Surveys with a non-bolded "x" in first column missed by a narrow margin).

Table 3. Dimensions of Teamwork Assessed by Surveys Used for Larger Work Groups[3]

				Behavioral Processes									Emergent States				
													Affective			Cognitive	
	Psychometric Validity	**Related to Outcomes**	General teamwork quality	Communication	Use of all contributors' expertise	Coordination (mutual adjustment)	Collaboration	Active conflict management	Effort	Shared decision making	Help each other/share workload	Respect	Social support	Psychological safety	Role responsibility understanding	Shared objectives
Adams 1995			■	■	■	■	■	■	■	■	■	■	■		■	
Kalisch 2010	**X**	X	■	■	■	■	■	■	■		■	■	■		■	■
Shortell 1991	X	X	■	■	■	■		■	■	■		■			■	■
Sorra/AHRQ 2004		X	■	■		■	■					■	■	■	■	■
Ushiro 2009			■	■	■	■	■	■	■		■	■	■		■	■
Baggs 1994		X	■	■	■	■	■			■					■	■
Gittell 2002		X	■	■	■			■				■			■	■
Copnell 2004			■	■	■	■	■			■					■	■
Sexton 2006		X	■	■	■	■	■	■				■	■	■	■	■
Masse 2008			■	■	■			■				■		■	■	■
Hutchinson 2006			■	■											■	■
VanBeuzekom 2007			■												■	■

[3] Surveys listed in rows, sorted by number of teamwork dimensions assessed. Teamwork dimensions listed in columns, sorted within categories by number of surveys by which each dimension was assessed. An X in the first column indicates that a survey met all criteria for psychometric validity (Table 4), and an X in the second column indicates that a survey has an established relationship with a non-self-reported outcome. (Surveys with a non-bolded "x" in first column missed by a narrow margin).

Table 4. Psychometric Properties of Survey Instruments that Measure Teamwork

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
	SCALES FROM TABLE ONE							
	Cross-functional Cooperation	Pinto 1993	Cross functional Cooperation scale, 15 items 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 1993	Full survey, 54 items, 3 items each in Communication/ cooperation within work group, Participation 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 1993)
	Group Effectiveness/ Interdisciplinary Collaboration	Vinokur-Kaplan 1995 /Armer 1978	Collaboration scale, 10 items 7 point scale	Not reported	Collaboration scale, 0.82	Based on previously validated and implemented scales (Armer 1978)	Not reported	Positively associated with -objective standards of quality met, team cohesion, and overall team effectiveness (Vinokur-Kaplan 1995)
	Team Process Domain	Denison 1996	Team Process, 21 items 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 1996)
X	Psychological Safety and Team Learning	Edmondson 1999	Psychological safety, 7 items Team learning behavior, 7 items 7 point scale	Intraclass correlation coefficients: Psychological safety, 0.39 Team learning behaviors, 0.33	Psychological 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 1999) -greater team engagement in quality improvement work (Nembhard 2006)

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
X	Team Effectiveness Audit Tool	Bateman 2002	Full survey, 46 items 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.	Two types of factor analysis (Cattell's scree test and eigenvalues >1) identified a four-factor solution FL>0.3 VarExp: 72.3%	Original paper develops and validates survey instrument
X	Team Process	Doolen 2003	Team Process, 5 items 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 ($p<.05$)	Positively associated with -self-report team effectiveness and satisfaction (Doolen 2003)
	Team Diagnostic Survey	Wageman 2005	Process criteria scale, 9 items Team social process, 7 items 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 (conservative test of DV) showed that the scales have weak DV (pg 391).	Positively associated with -objectively measured team performance (Wagemen 2001) -team effectiveness (Hackman and O'Connor 2005)
	Team Survey	Senior 2007	Full survey, 36 items 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 1998	Exploratory factor analysis found a 5 factor solution, with one factor including 3 sub-factors. FL > 0.40 VarExp: 54%	Original paper develops and validates survey instrument
	SCALES FROM TABLE TWO							
	Team Process Scale	Brannick 1993	Team Process scale, Number of items not reported Response scale not reported	Rwg 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 (i.e. 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 lab Cited in health care simulation studies of teamwork

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
	Team Member Exchange (TMX) Quality Scale	Seers 1995	TMX scale, 10 items 5 point scale	Not reported	TMX scale, 0.83	Based on Seers' earlier TMX scale, developed for individual level of analysis	Not reported	Gains in departmental efficiency related to average change in scale over time (Seers 1995)
	Collaboration Scale	Kahn 1997	Collaboration scale, 6 item 5 point scale	Not reported	Collaboration scale, 0.92	Scale is based on literature/previous studies	Factor analysis revealed a uni-dimensional 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 1998	Full survey, 38 items 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 2003) -patient satisfaction (Proudfoot 2007) -quality of work in medical labs (Pitt 2002) -lower turnover in health care teams (Kivimaki 2007)
	Team Process Quality	Hauptman 1999	Team Process Quality Scale, 16 items, 5 pt. ordinal scale	Not reported	Team Process Quality Scale, 0.75-0.77	Questionnaire was pre-tested through semi-structured interviews with managers involved in NPD 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 2001	Full survey, 40 items 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 four 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

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
X	Team Effectiveness	Pearce 2002	Team Effectiveness, 26 items 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 (1992), Manz and Sims (1987), and Cox (1994)	Factor analysis revealed a uni-dimensional construct for effectiveness	Team effectiveness is the outcome variable (Vertical and shared leadership are predictive of greater team effectiveness)
	Team Functioning	Strasser 2002	Team Relations, 45 items Team Actions 27 items True/False 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 2005	Team participation, 7 items Team functioning, 8 items 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 two distinct factors as hypothesized.	Team participation associated with improvements in patient functioning, Team functioning was not significantly associated with patient functioning: (Alexander 2005)
X	Teamwork Quality Survey	Hoegl 2001	Teamwork scale, 37 items 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 2006)
	Teamwork Scale	Friesen 2008	Teamwork scale, 9 items 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 2008)

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
	Team Organization	La Duckers 2008	Team organization, 5 items 7 point Likert scale	Not reported	Team organization, 0.84	Development included two phases: first a literature review and expert assessment of the clarity, completeness of questions; and pilot test to determine psychometrics	Principal component analysis revealed 3 factors FL > 0.5 VarExp: 15%	Original paper develops and validates survey instrument
	SCALES FROM TABLE THREE							
	ICU Nurse Physician Collaboration	Shortell 1991	Full survey, 82 items Coordination scale, 13 items Communication scale, 43 items Problem-solving scale, 14 items 5 point Likert scale	Tested using ANOVA: variance within the units significantly less than variance between units ($p<.05$)	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 semi-structured 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 1994) -lower incidence of mortality and chronic, severe morbidity in NICU (Pollack 2003)
	Collaboration and Satisfaction about Care Decisions	Baggs 1994	Collaboration scale, 7 items 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 1999) -nurse satisfaction with decision making (De Chairo 2001)
	Professional Working Relationships	Adams 1995	Professional Working Relationships, 26 items 4 point response scale	Not reported	Professional Working Relationships, 0.84-0.91	Extensively tested through qualitative work, literature review, pre-test, 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 2004)

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
	Relational Coordination	Gittell 2002	Relational coordination scale, 28 items (7 items relating to 4 other disciplines) 5 point Likert scale	Cross-group differences in relational coordination tested using ANOVA ($p<0.01$)	Relational coordination scale, 0.8	Based on previously validated scale (Gittell 2000)	Not reported	Positively associated with -quality of care, postoperative functioning; negatively associated with postoperative pain and length of stay (Gittell 2000) -patient functional status, mental health, and freedom from pain (Gittell 2002)
	Hospital Survey on Patient Safety	AHRQ 2004	Full survey, 42 items Teamwork within units scale, 4 items Organizational learning scale, 3 items Communication openness scale, 3 items 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.	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 2010) Further validated in Sorra (2010)
	Perceptions about Interdisciplinary collaboration scale	Copnell 2004	Full survey, 29 items 5 point Likert scale	Not reported	Not reported	Adapted from Anderson (1996), 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 2006	Teamwork scale, 22 items 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

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
	Safety Attitudes Questionnaire	Sexton 2006	Full survey, 40 items Teamwork climate scale, 6 items 5 point Likert scale	Not reported	Full survey, Raykov's coefficient: 0.90	Literature review to develop items, pilot tests and revisions of items and structure	CFA confirmed hypothesized six factor structure, Teamwork scale, FL: 0.76-0.96	Communication and collaboration were associated with lower risk-adjusted morbidity, not associated with mortality (Davenport 2007) Scores improved following an intervention (Sexton 2011)
	Leiden Operating Theater and Intensive Care Safety (LOTICS)	Van Beuzekom 2007	LOTICS, 40 items 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 2008	Collaboration scale, 23 items 5 point Likert type response	Not reported	Collaboration scale, 0.75-0.91	Questions developed based on pre-existing conceptual models (Rosenfeld 1992) 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 2009	Collaboration Scale, 27 items 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 three factors. The three-factor model was confirmed by confirmatory factor analysis. FL > 0.4	Negatively related to -nurses gender role attitudes (Ushiro 2010)

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained

Table 4: Continued

X	*Scale*	*Source*	*Number of items, Response scale*	*Inter-rater agreement and reliability* [a]	*Internal consistency/ reliability*[b]	*Content validity*	*Discriminant validity*[c]	*Validated relationships to outcomes of interest*
X	Nursing Teamwork Survey	Kalisch 2010	Teamwork Survey, 33 items 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 2005). Focus groups conducted to develop items within categories. Experts reviewed each questions and suggested modifications or elimination.	Exploratory factor analysis yielded five factors. The five-factor model was confirmed by confirmatory factor analysis. FL > 0.4	Positively related to -higher staffing levels (Kalisch 2011) -job satisfaction (Kalisch 2010) -missed nursing care (Kalisch 2012)
SCALES MEASURING ATTITUDES TOWARDS TEAMWORK								
	Attitudes towards Health Care Teams	Heinemann 1999	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 1999	Full survey, 20 items 4 point Likert scale	Not reported	Full survey, 0.84	No qualitative or pilot testing reported.	Factor analysis generated four factors. FL > 0.40	Original paper develops and validates survey instrument Scale later used as outcome variable

a Value reported is rwg statistic unless otherwise indicated.
b Value reported is Cronbach's alpha unless otherwise indicated.
c PCA = principal component analysis, FL = factor loadings, EV = eigenvalues, CFA = confirmatory factor analysis, VarExp = Variance Explained