



**Center for  
Effective  
Organizations**

---

**TEAM EFFECTIVENESS IN  
MULTINATIONAL ORGANIZATIONS:  
EVALUATIONS ACROSS CONTEXTS**

**CEO PUBLICATION  
T 01-8 (398)**

**CRISTINA B. GIBSON**  
*Center for Effective Organizations*

**MARY E. ZELLMER-BRUHN**  
*University of Minnesota*

**DONALD P. SCHWAB**  
*University of Wisconsin, Madison*

**January 2001**

This research was made possible with funding provided by the NSF Grant #SBR 96-31748, the Carnegie Bosch Institute for Applied International Management Research, and the University of Wisconsin Initiative for World Affairs and the Global Economy. The researchers would like to acknowledge the time and effort extended by all of the respondents in this research together with their associated staff. Special thanks to Rauol Zapata in Puerto Rico, Joylie Agustin and Ricardo Lim at the Asian Institute for Management in the Philippines, and Michael Segalla at the Hautes Etudes Commerciales in France. We would also like to express our appreciation for the administrative support, translations, and transcriptions provided by Paula Bassoff, Ryan Billingham, Florence Brunell, Joan Donovan, Kerry Jung, Francisco Lloveras, Rachel Ritterbausch, David Robinson, Carol Troyer-Shank, Agnes Valley, and Richard Zapata at the University of Wisconsin.

## Abstract

Incorporating team context into research and practice concerning team effectiveness in multinational organizations is an on-going challenge. We argue that a common measure of team effectiveness with demonstrated equivalence across contexts expands current theoretical developments and addresses team implementation needs. To this end, we describe methods, techniques, and results obtained in developing a comprehensive team effectiveness survey across 6 multinational organizations in 4 geographic regions. We discuss inductively deriving dimensions of effectiveness using interview data; translation procedures with innovative bilingual pilot testing; and multiple constituency validation. We conclude with implications for future cross-cultural research on team effectiveness and beyond in other areas of international management.

## Team Effectiveness in Multinational Organizations: Evaluation Across Contexts

Use of work teams, groups of employees with interdependent interaction and mutually shared responsibility (Sundstrom, DeMeuse & Futrell, 1990), has increased dramatically during the past decade. Research conducted in the early part of the decade (Wellins, Byham & Wilson, 1991) suggested that only about one-quarter of the organizations surveyed were using teams, involving only a small portion of the workforce. Later, Osterman (1994) found that over 50 percent of the 700 organizational units he studied were using teams and that over 40 percent had more than half of their employees working in teams. Additional evidence suggests that this trend continues to gain momentum. For example, 60 percent of the 313 organizations investigated by Lawler, Mohrman, and Ledford (1995) claimed that they will increase or greatly increase their use of teams over the next decade; only 3 percent said they would reduce or discontinue the use of teams. Despite the increase in use of teams, measurement of individual performance is still the focus of most research and many appraisal and reward systems (Mohrman, Cohen & Mohrman, 1995). When measured, team level outcomes are often nested in departmental or unit-wide measures. It is often impossible for researchers and managers to compare teams in different functional areas, departments, or facilities (Cohen & Bailey, 1997). This makes both the investigation and design of performance review, feedback, and reward systems difficult. Research and practice regarding the reinforcement of specific and desired team behaviors is a particular challenge.

To complicate matters, many organizations now have operations across multiple nations. For example, a decade ago, already one third of the profit in U.S. companies was derived from international business, along with a sixth of the nation's jobs (Cascio, 1989). Recent estimates

indicate that internationalization is at least twice as prevalent now (Earley & Erez, 1997). While a number of programs of research have investigated the implementation of work teams in a domestic U.S. environment, a key facet that has been ignored are the special concerns that face multinational organizations as they design team-based systems in facilities located in different nations. Practitioners in these organizations struggle with the assessment of teams across the cultures that represent these national contexts. Many team-oriented phenomena are culture-specific (Gibson, 1999) and the acceptability of a team-based system varies across cultures (Kirkman & Shapiro, 1997). As a result, research and development of equitable performance review, feedback, and reward systems for teams across the various facilities of a multinational organization become problematic.

Our review of the team effectiveness literature suggests we are now at a key juncture in theory development. To truly push the field forward, we argue that a more explicit consideration of team context is critical, with a particular emphasis on the cultural context. To facilitate this, we further argue for the importance of a measure of team effectiveness that is applicable across teams, organizational contexts, and cultures. Such a measure is essential to both empirical research on theoretical models designed to understand determinants of team effectiveness and to the implementation of teams in multinationals. This is particularly true given the emerging centrality of contextual factors associated with the organization and cultural settings in which teams are embedded. As we demonstrate, developing such a measure will substantially aid in the extension of various team theories, such as task design, leadership and motivation, to multinational settings.

In addressing this need, our paper unfolds as follows. First, we review the literature on teams and propose a shift in theoretical developments made possible by team effectiveness

measures that generalize across contexts. Next, we describe our review of existing measures, develop an interview protocol, and derive potential items for the measures from exploratory cross-cultural interviews. We then discuss the process used to refine these items and create initial instrumentation. We also describe the techniques we used to ensure equivalence of meaning across foreign language versions of the measures. Next, we present a comprehensive bilingual analysis of the instrumentation, and review the results of exploratory bilingual tests. Finally, we describe a multiple constituency approach to pilot test our measures in the field, and report results of an exploratory quantitative analyses of the pilot study data. We conclude by discussing the implications of our research.

### Theoretical Background

Coinciding with the increased use of teams by practitioners, the research literature on teams has a rich and varied history (see Cohen & Bailey, 1997 for a review). Critical research has investigated, for example, the nexus of teams and task design theory (Campion, Medsker & Higgs, 1993), leadership theory (Cordery, Mueller, & Smith, 1991), motivation theory (Wageman, 1995), and learning theories (Ancona & Caldwell, 1992b). Empirical studies investigating these theories have typically included team outcome measures that are organization-specific or unit-specific, such as a particular type of error rate for manufacturing teams (Wall, Kemp, Jackson, & Clegg, 1986), a complex customer satisfaction index for customer service teams (Wageman, 1995), or a specific calculation of time-to-market for product development teams (Ancona & Caldwell, 1992b).

In addition, researchers have recognized the importance of the general context in which teams are embedded and relationships between teams and their external environment (e.g.

Ancona, 1990; Ancona & Caldwell, 1992a; 1992b; Cohen & Bailey, 1997; Gladstein, 1984).

These researchers acknowledge that teams are not socially isolated entities and that both internal and external processes have important implications for team performance. As multinational organizations develop complicated team-based structures across their national facilities, organizational and cross-cultural factors have become additional aspects of team context that need to be incorporated into theory development about teams.

To move the field forward in terms of both theoretical development and multinational management, we argue that theories of team effectiveness must be expanded to better accommodate variations stemming from complex contexts in which teams are embedded. Furthermore, we argue that accomplishing this requires an approach to measuring team effectiveness that is generalizable across contexts. For example, task design theory explores various characteristics of jobs, such as autonomy or variety, and the impact of these characteristics on outcomes such as motivation and satisfaction (Hackman, 1990; Jehn, 1995). To address contextual features and expand upon task design theory for teams, a key research question might be, “Are job characteristics related to team effort across different organization contexts?” Potentially, the most robust investigation of this issue requires including teams from multiple organizations in the same study, and a team effectiveness measure that is applicable across these organizations. As of yet, no such instrument has been comprehensively developed, potentially thwarting theoretical development of task design theory for teams that is applicable across organizational contexts.

As a second example, to incorporate contextual features and expand upon leadership theory for teams, a key research question is, “Are external leadership behaviors related to team effectiveness across organizations?” Rigorous investigation of this question requires that

multiple organizations be included in a project, and measures must be applicable across varying organizational contexts. Until this research has been conducting, the generalizability of theories pertaining to the impact of leadership behaviors on team effectiveness is uncertain.

In a similar vein, to acknowledge contextual features and expand upon motivation theory for teams, a key research question becomes, “Are group rewards motivating across cultures?” To research this issue, teams in the same organization located in multiple cultural contexts must be included in the research, and the measure of team outcomes utilized must be applicable across these settings. Until we have tested these relationships in multiple cultures, we cannot fully develop theory regarding the motivational efficacy of rewards for team effectiveness in different national contexts.

As a final example, given the increased focus on organization and cultural context, to expand upon learning theory for teams, a key research question becomes, “Can teams transfer knowledge across geographic and organizational boundaries?” The most systematic investigation of this phenomenon requires that teams in multiple organizations located in multiple cultural contexts be included in a sample. Furthermore, measures of team outcomes must be applicable across varying organizational contexts and consistent in meaning across languages and cultures. Without such a measure, it is difficult to establish the generalizability of a theory of team learning and knowledge transfer across multiple boundaries.

Conducting any of this research on teams across such contexts requires a shift in theoretical framing from an emic (culture-specific) to an etic (culturally general) framework. Much of the research conducted in the field of international and cross-cultural management reflects a basic disagreement among academics about the relative importance of each of these approaches. This debate queries, “What is the appropriate level of analysis for understanding

cultural influences on work activities?” Most research is emic, conforming to what Earley and Singh (1995) referred to as a “reduced” form of research design, meaning that culture is treated as one of many exogenous factors that influence management processes. For example, one might argue that people from a high power distance culture are more likely to respect highly autocratic leaders. This type of argument reflects cultural factors isolated from one another and in relation to a very specific outcome.

Earley and Singh (1995) argued that a superior form of research design was reflected by a “hybrid” design, incorporating several features of culture. First, in developing research questions, gestalt systems are studied in order to identify important aspects of the systems. Second, hypothesized relationships are derived across systems and they are not necessarily unique to a given system. Third, constructs and relationships are assumed to be separable from the system in which they are embedded, but the application of the constructs and relationships in each context may not be exactly the same. Fourth, specific relationships are interpreted using reduced parts of the system, but with reference to the general system. These interpretations can, in turn, lead to a further refinement of general principles.

An interesting example of this style of research is Van Maanen and Barley’s (1984) assessment of occupational communities. The inferences they draw from prior empirical work illustrate the complementary mixing of ethnographic observation, conceptualizations derived from within a given system extended onto other systems, and the development of a general theoretical model that captures general principles across many systems. For instance, they examine a number of occupational systems through an intensive immersion to understand why individuals within the system behave as they do, and the nature of each community’s social structure. Next, they take these observations and combine them in order to understand the

patterns of actions across multiple communities using several theoretical perspectives as a means of discussing these observations. We argue that a similar approach is necessary to expand upon current team-focused theory. The purpose of this paper is to describe one such measurement development effort.

We used a multifaceted approach to achieve this purpose. With prior research on team effectiveness as a guide, we sampled across organizations and cultures, relying largely on qualitative methods to inductively derive measures and confirm cross-cultural validity. Overall, the development and evaluation was conducted over a 12-month period; it involved site visits to six organizations that use teams for a variety of work activities. These organizations all operate internationally and teams from each of the six organizations were visited in four countries, for a total of twenty-four sites. Fifteen translators and analysts used a systematic combination of procedures to foster cultural equivalence among the items developed to measure team outcomes. These procedures included review of existing measures, bilingual interviewing and transcription, item generation and review, item translation and back translation, pilot testing, and finally exploratory psychometric analysis. Our emphasis was on equivalence of meaning, and our primary method of investigating equivalence was using qualitative methods supplemented by exploratory quantitative analysis.

Second, we wanted to create an instrument that would be useful to both researchers and practicing managers in multinational organizations. Our observations suggested that many organizations are using “360 degree” feedback systems to measure individual performance. In the research literature, such systems have often been referred to as multiple constituency ratings (e.g., Gladstein, 1984; Tsui, 1984). At the individual level, such a technique involves obtaining ratings from a target individual about their own performance, from the target’s supervisor, and

from his or her “customers.” Customers can be either internal or external to the organization, and are often described as the party downstream in the process who receives the work of the target individual. The technique of gathering multiple constituency data was developed primarily to avoid social desirability response bias, in which individuals inflate ratings of their own behavior (Nunnally, 1978). Generalizing to the team level of analysis, it seems equally useful to obtain ratings of a team from multiple constituents, all of whom could potentially provide useful information and feedback about a team (Tsui, 1984). Previous research has included ratings from multiple raters (e.g. Ancona & Caldwell, 1992a; Cohen & Ledford, 1994; Cohen, Ledford & Spreitzer, 1996; Jehn, 1995; Jehn, Northcraft, & Neale, 1999), and many of our interview respondents concurred, commenting on the need to obtain “360 degree” assessments for team effectiveness. This need arises particularly because teams may inflate self-ratings of team effectiveness, just as individuals often inflate ratings of their own performance. We examined this potential bias by creating different survey formats for team members and team leaders, and we extend the work of previous researchers by creating a third format for the ratings by team customers. We explore the reliability of these of forms across rater types. We elaborate upon this process in the next section.

## Methods

### Derivation of Items

The measurement effort described here grew out of a four-year program of research designed to investigate the utilization of teams in multinational organizations, using multiple methods across cultures. A key component of this research involved the design of measures applicable across organizations and cultures. We began this effort by conducting an extensive

review of existing measures of team effectiveness that appeared in the key managerial and social psychological journals (see Appendix A) over the last two decades. For example, we drew extensively from the work of Cohen and her colleagues (Cohen & Ledford, 1994; Cohen, Ledford, & Spreitzer, 1996; Cohen & Bailey, 1997), from the work of Hackman and colleagues (Hackman, 1986; 1990; Hackman & Oldham, 1975), and from the work of Ancona and her colleagues (Ancona, 1990; Ancona & Caldwell, 1992a; 1992b). We utilized this literature to develop our interview protocol (see Appendix B), to inform our sample selection, and to conduct the qualitative analysis for item generation (see below).

Based on previous research, we selected four regions of the world - U.S., France, Puerto Rico, and the Philippines - to obtain variance on two key cultural characteristics, power distance and collectivism, that demonstrate important implications for teamwork (Earley & Gibson, 1998; Hofstede, 1980; Kluckhohn & Strodtbeck, 1961; Maznevski & DiStefano, 1997). Power distance represents the degree to which members of a culture accept and expect that power in society is distributed unequally (Hofstede, 1980). Cultures low in power distance will try to minimize inequalities, favor less autocratic leadership, and favor less centralization of authority. Research suggests that the U.S., Australia, Canada, Denmark, and Austria are low in power distance (Hofstede, 1980). On the other hand, cultures high in power distance will be characterized by greater acceptance of inequalities, more autocratic leadership, and greater centralization of authority. Research suggests that Malaysia, the Philippines, Panama, Puerto Rico and France are relatively high in power distance when contrasted with the previously mentioned nations (Hofstede, 1980).

A second key cultural value pertinent to the management of teams is collectivism. Collectivism describes the strength of ties between individuals in a society, the degree to which

members are integrated into groups, and the extent to which members of a society value their membership in groups (Hofstede, 1980). Countries such as Peru, Chile, Puerto Rico, Taiwan, Singapore, and the Philippines are highly collectivistic; countries such as the U.S., Australia, New Zealand, Italy, and France are low on collectivism.

Members of teams were interviewed in six major multinational organizations that produce, sell, distribute and service diversified products that cross industry classifications. In each firm, representatives from teams located in each of the four geographical regions were interviewed. Human resource professionals in each multinational firm were asked to select three to four teams per facility for the interviews. We asked that they identify a variety of functional team types, from a variety of levels in the organization, and from teams that were both poor performers and high performers.

The final interview sample included 59 teams represented by 126 individuals. We interviewed an average of two to three persons per team. Team size varied from approximately five to ten members. A variety of teams were interviewed to increase the generalizability of our instruments, including manufacturing teams, customer service teams, sales and marketing teams, administrative teams, management teams, and project teams.

The researchers traveled to each region and met with each interviewee face-to-face for approximately one hour. A pre-established series of questions were posed pertaining to the function of the team, teamwork knowledge, the management of the team, and the context in which the team works (see Appendix B). Interviewees were also asked to discuss factors they felt were important facilitators and inhibitors of team effectiveness. Approximately 80% of the interviews were conducted in English; however, a bilingual interpreter was present during interviews in non-U.S. locations to assist in any necessary translations. We taped-recorded all

interviews and had the tapes transcribed by bilingual professional transcriptionists. This served as a check on the adequacy of any original translation that took place during the interviews.

Whenever a word or phrase was not directly translatable, it was retained in the original foreign language. We used the transcriptions to create a database consisting of over 1,000 pages of single-spaced text.

Working in an inductive mode, we content-analyzed the interviews using computer facilitated qualitative data analysis. We derived items for team effectiveness using an eight stage process: (1) creating key word list, (2) excerpting text containing key words, (3) deriving items, (4) deriving dimensions, (5) coding items and assessing inter-rater reliability, (6) compilation of dimensions, (7) item revision, and (8) modification of items for multiple constituents. Each of these stages is discussed in turn below.

Stage 1 - Key word list. We identified a set of key words using several review articles in the teams literature pertaining to team effectiveness (Cohen & Bailey, 1997; Goodman, Ravlin & Schminke, 1990; Shea & Guzzo, 1987; Sundstrom, DeMuese & Futrell, 1990). The terms included were: performance; evaluate; evaluation; effective; effectiveness; internal customer; customer; measure; measured; measurement; satisfaction. Next, we used an electronic search function to search for each of these terms in the 1,000-page database. Any time a term was found, it was highlighted with a boldfaced color and underlined, regardless of whether the interviewee was responding specifically to questions about effectiveness, or to other questions. This procedure provided a rich understanding of how our interviewees viewed effectiveness because it captured both "evoked" and "natural" data on team effectiveness. The source of evoked data is direct and typically transparent questioning; alternatively, natural data comes from sources in which the subjects have no way of knowing how the text data will be used (Kabanoff, 1997). In

our questions directly addressing effectiveness, the interviewees were conscious of our interests. In other questions, they discussed additional features of effectiveness that may not have occurred to them in direct questioning. Thus, using key word searches rather than simply the answers to the direct questions on effectiveness provided us with a broad potential set of items and dimensions.

Stage 2 - Excerpt files. Once all key terms were highlighted in the transcript we visually scanned the transcripts to determine where the discussion related to the term began and where it ended. This excerpt was then copied and pasted into a new document. This process resulted in four separate "excerpt files," each representing responses from a different country. We retained the codes for organization in these files to allow for organizational level as well as national level analyses. The excerpt files contained all segments of text in which any outcome effectiveness term was mentioned.

Stage 3 - Deriving items. We carefully examined the excerpt files to determine how each interviewee defined team outcome effectiveness. Each expression of team effectiveness that had not been previously encountered in the transcripts was considered a "unique item." For example, in one interview transcript, there was a 30 - 40 line excerpt surrounding the occurrence of the word "effectiveness." When this exchange was examined, it was clear that for this interviewee, effectiveness was defined as "number of errors made." Number of errors became one unique effectiveness item. Table 1 summarizes the 24 unique items for team effectiveness obtained from this procedure. We also noted the organization and country in which the items were mentioned. This allowed us to track how frequently items were mentioned across countries and organizations.

-----  
Insert Table 1 about here  
-----

Stage 4 - Deriving dimensions. Two of the researchers independently reviewed the excerpts from Table 1 to identify whether the items could be grouped together into dimensions. In accordance with the principal of parsimony, each researcher attempted to identify the least number of dimensions possible without losing uniqueness of the items. The two researchers then compared their independently created lists of dimensions. The researchers both independently identified eight dimensions of outcome effectiveness: goals, stakeholder satisfaction, timeliness, errors, productivity, change, shared mission, and market share.

Stage 5 - Coding items and assessing inter-rater reliability. After identifying dimensions, the researchers categorized each item derived from the interviews into a dimension. Their inter-rater agreement on the categorization (Cohen's Kappa) was .65. To resolve disagreements, the researchers consulted literature regarding constructs that were similar to the dimensions. For example, an interviewee suggested that "number of days to complete a task" was an indicator of team effectiveness. One researcher categorized this item in a "productivity" dimension. The second researcher categorized this item in a separate dimension representing "timeliness." In resolving this discrepancy, the researchers consulted Daft's work which defines productivity as the "organization's output of products and services divided by the inputs" (Daft, 1997: 729). The item mentioned by the interviewee (number of days to complete a task) refers to inputs in production process (e.g., time), but does not mention the nature of the outputs provided; therefore, the item did not appear to coincide with the concept of "productivity." As a result, the item was categorized in accordance with the interpretation made by the second researcher (in a dimension called timeliness). Table 2 describes the dimensions and summarizes the items

grouped under each dimension.

-----  
 Insert Table 2 about here  
 -----

Stage 6 - Compilation of dimensions. After resolving the categorization of items into dimensions, the frequency that each dimension was mentioned across countries and organizations was computed. The dimensions discussed most frequently by respondents were also those that occurred most broadly in the greatest number of countries and organizations, providing support for the generalizability of the dimensions. Any dimension that did not occur frequently or broadly across organizations and countries was collapsed into other dimensions. As a result, the dimension, “market share” was removed from subsequent analysis because it contained only one item and appeared to be related to a narrow set of team types (e.g., sales). The single item under “mission” was included with the goals dimension. Finally, the three items in the change dimension were revised and included in the productivity and quality dimensions. Through these procedures a set of five generalizable outcome dimensions were delineated: goals, customers, timeliness, quality, and productivity. The goals dimension represents the degree to which the team meets its objectives. The customers dimension is the extent to which the team addresses and meets customer needs. Timeliness is the degree to which the team is efficient with respect to time and adheres to temporal goals. The quality dimension is the extent to which the team produces error free work. Finally, productivity is the degree to which the team is efficient with respect to the ratio of inputs to outputs.

Stage 7 - Item revision. The items identified from the interview excerpts often contained broad statements that encompassed more than one complete thought. Other items were too brief to be clear. Thus, the original set of items were revised so that each contained one complete

thought. In addition, we reworded some items to be reverse scored. As a result, for each dimension, six to eight items were developed. Table 3 provides a summary of the dimensions and items that were derived from these steps.

-----  
 Insert Table 3 about here  
 -----

Stage 8 - Modifying items for leader and customer instruments. Our final stage in item development was to create formats that could be used by different raters. We created three different formats of the survey instrument: member, leader, and customer. The team member survey asks members to rate their own team, and contains statements such as “Our team accomplishes its goals.” Items were modified for the leader and customer surveys, which contain statements such as “This team accomplishes its goals.” Creating three formats allowed us to test convergent validity across raters.

#### Translation and Back-Translation

Upon arriving at a set of items, we used a comprehensive procedure to translate the items into French and Spanish. First we identified bilingual speakers to conduct the initial translation from English to French and from English to Spanish. We selected a native Puerto Rican and a native Parisian to reflect the specific regions involved in the project. We also screened for work experience to ensure that the translators were familiar with business terminology and the constructs used on the survey.

We instructed the translators to translate from the English version of the survey all items, scales and instructions. Translators brought any terms that had no direct translations to our attention. In these instances, the researchers identified a different English term that could be accurately translated; there were fewer than five such changes in the initial translation. The

translated versions of the survey were then subjected to back-translation, which serves as a check on the initial translation (Brislin, Lonner, & Thorndike, 1967). We selected a second set of translators for the back-translation using the same procedures described above. When the back-translators had completed their work, two of the researchers reviewed the original English survey and the back-translated English to ascertain whether the two versions were substantially different in meaning. Items that appeared to carry different meaning in the back-translated English were highlighted for further examination.

To investigate these items, we conducted a focus group including the researchers and four bilingual translators, one of who was fluent in all of the languages used in the study (English, French, and Spanish). During the focus group, we collectively reviewed the problematic items on the original English surveys and the back-translated surveys to identify whether: (1) the original English could not be readily translated, (2) the translation could be improved, or (3) the translation back to English could be improved and would then result in similar meaning to the original. Problem type (1) was resolved as the researchers conferred with the translators to change the English items in such a way that the meaning of the item was retained but the words could be more readily and accurately translated into both French and Spanish. Problem type (2) and (3) were resolved by having the translators reach a conclusion about a better translation. We compiled these resolutions and made changes in all three versions of the surveys (English, French, and Spanish) to reflect these decisions.

### Pilot Testing

We designed two studies to assess reliability and conduct validation of the a priori scales: (1) a bilingual pilot test and (2) a team-level, multiple-constituent pilot test. The bilingual pilot test was designed to examine the reliability across language versions and to supplement overall

scale validation efforts. The team-level pilot test was used for overall reliability tests and scale validation, and to examine the performance of the instruments using multiple constituents; more specifically, to ensure that scales adequately capture ratings from team members, team leaders and team customers.

Bilingual pilot study design. In addition to the back-translation process, Brislin, Lonner, and Thorndike (1967) recommend that translated surveys be pilot tested in each language to further examine validity of the items in the translated versions. The bilingual pilot test was conducted at the individual level of analysis. It was designed to ascertain whether bilingual speakers answered the items consistently, regardless of the language presented. We wanted to guarantee that respondents did not infer a different meaning for the same question across languages. In addition to testing for consistency of meaning, we also used the results from this study to examine overall reliability of the scales.

Bilingual Spanish/English speakers and bilingual French/English speakers were identified among the graduate student body on a large university campus and among the working population in a Midwest community through mailings, announcements in classes, and posters. Participants were paid \$10 to participate. The sample included a total of 18 French/English and 18 Spanish/English bilingual respondents. Respondents were told to concentrate on a team they were currently serving on, or had just recently served on, when completing the survey. They completed both surveys consecutively during a two-hour period. When they had completed the first survey, they were instructed to wait approximately one hour before completing the second version. The authors observed the participants and ensured that they did not copy answers from one version to another. Survey order was randomly counter-balanced to avoid any systematic response bias that may occur from completing the questionnaire twice. Half the respondents

completed the English version first, the other half completed the Spanish or French version first.

In addition to completing the survey in both languages, the bilingual pilot participants served as one final check on the translations. They were instructed to make notes in the margins of the instrument if any of the items were confusing, or if the language or grammar seemed inappropriate. This feedback was incorporated into the final versions of the scales.

Multiple constituency pilot test design. The second pilot test examined the performance of the items at the group level of analysis and across multiple raters (team members, leaders, and customers). Four teams from a small, biotech firm, five teams from a large, multi-national pharmaceutical organization, and two teams from a large HMO completed the survey. External constituents for each team (leaders and customers) also completed the survey. The total sample size for the multiple-rater pilot was 11 teams rated by 99 individual raters. Each team had at least one external rater.

### Analyses and Results

We performed several sets of analyses to test the survey items and scales. These analyses included internal reliability analysis (coefficient alpha), an examination of item-to-total correlations, and multiple regression. To aid in the analysis, we created a total of seven databases using the various pilot data: (1) English surveys completed by French/English bilinguals, (2) English surveys completed by Spanish/English bilinguals, (3) French surveys completed by French/English bilinguals, (4) Spanish surveys completed by Spanish/ English bilinguals, (5) team members in the multiple-constituent pilot study, (6) leaders in the multiple-constituent pilot study, and (7) customers in the multiple-constituent pilot study.

Reliability and item dimensionality. The first set of analyses was performed separately on each database to allow comparison of internal consistency and item-to-total correlations across

types of raters (leader, member, and customer) and across languages (English, French and Spanish). First, we conducted internal consistency reliability analyses for each of the five a priori effectiveness dimensions and computed item-to-total correlations. We then developed Table 4 containing alpha for each scale and identifying any item that could be dropped to improve the scale reliability. Developing this table allowed comparison of reliabilities across all the databases to determine if any of the scales were problematic in a given form (i.e., leader, member, customer, French, English, Spanish). Items that reduced the reliability across most or all of the databases were removed. Items that had low item-to-total correlations in the French or Spanish versions, but not in the English version, were identified as potential translation problems and subjected again to the translation-back-translation process.

-----  
 Insert Table 4 about here  
 -----

Creation of scale scores. After modifications to the scales were made, we computed scale scores for each of the remaining dimensions. These scores were created at the individual level by computing the average for each person across all items in a scale. We used these scale scores in the remaining analyses.

Bilingual agreement across language versions. Several regression analyses were performed to examine consistency within a respondent across language versions. The purpose of these analyses was to determine if the mean score in English was significantly different from the mean scores in the two translated versions. We conducted these analyses using scale scores for each of the effectiveness dimensions. Each scale score served as a dependent variable in five separate regressions. Because there are three different versions of the survey (English, Spanish & French) we created *k-1* dummy coded variables (two variables - one for scales answered in

Spanish and one for scales answered in French). The omitted category was the English version. In dummy-coded regression, with no other independent variables, the regression coefficients can be interpreted as the difference between the mean score of the dummy coded group and the mean score in the omitted category. For example, the Spanish partial regression coefficient indicates how much more (or less) the mean Spanish scale score is from the mean English scale score. Thus, the coefficients for each language version should be small if the translation procedure was effective.

Assessment of discrimination and convergence. We next performed a set of analyses to investigate the sensitivity of the five scales to role (member, leader, and customer) and team differences. Our objective was to measure team effectiveness independent of the role of the individuals providing the evaluations. Thus we sought differences between teams on scales, but no differences between rater roles. Data for the eleven teams provided by 99 raters described above were used for these analyses. Each of the teams had raters representing all three roles.

We conducted assessment of discrimination and convergence using multiple regression to analyze the percentage of scale variance explained by teams and roles. Dummy variables were created for  $k - 1$  teams (ten teams) and a regression analysis was run for each scale, with the scale score regressed on the team dummy variables. These regressions demonstrated whether the scales can be used to differentiate between the behavior of teams; large coefficients of determination are desired. Next, we created dummy variables for  $k - 1$  roles (two variables – members were the omitted category) and computed the regressions adding the dummy variables for roles to the team dummy variables. The expectation was that the addition of variables for role would not appreciably increase scale variance explained.

Results. Final scale reliability scores (Cronbach's alpha) for each of the outcome

effectiveness scales ranged as follows across the databases: goals (.73-.92), customers (.81-.96), timeliness (.62-.88), quality (.62-.88), and productivity (.46-.81). Table 5 summarizes the reliability scores for each scale in each database.

-----  
 Insert Tables 5 & 6 about here  
 -----

Results of the bilingual pilot regression analyses are summarized in Table 6. For each scale, we show the partial regression coefficients for the Spanish and French versions (and the corresponding t statistic values) and the multiple coefficient of determination adjusted for degrees of freedom (and corresponding F values). As expected, results indicate that language version had little impact on mean scale scores. On the productivity scale, language version accounted for two percent of the variance; however, this was not statistically significant. The negligible size of the partial correlation coefficients and coefficients of determination suggest that average scale scores are about the same regardless of language version.

-----  
 Insert Table 7 About Here  
 -----

Table 7 summarizes results from the multiple rater regression analyses. The left side of the table shows results relevant to discrimination between teams. The first two columns present the coefficients of determination (adjusted for degrees of freedom) and the F statistics from the regressions for teams. As expected, all equations are statistically significant, indicating significant between-team variance. Adjusted variance explained by teams ranges from a low of 10 percent for the customer satisfaction scale to a high of 22 percent for the quality scale. Characterizations of team behaviors differ systematically across teams on all of the scales. The right side of Table 7 shows results relevant to an assessment of convergence across roles.

Specifically, we examined the importance of roles in accounting for scale score variance. The partial regression coefficients and their associated t statistics demonstrate that customer and leader evaluations do not differ significantly ( $p < .05$ ) from member evaluations for any of the scales. For three of the five scales the addition of the two role variables resulted in no change or in a decrease in the coefficients of determination (adjusted for degrees of freedom). Furthermore, the non-significant incremental F values ( $p < .05$ ) indicate that, in combination, roles did not add appreciably to the variance explained for any of the scales. Thus, in contrast to individual self-assessments where there is evidence of leniency bias (e.g., Nunnally, 1978), we found little evidence that teams inflate self-ratings.

-----  
 Insert Table 8 about here  
 -----

These analyses thus found little evidence that results are influenced by language version or participant role. However, scores are influenced by differences among teams. Given these differences, we aggregated the data to the level of the team ( $N=11$  teams). Table 8 summarizes descriptive statistics (means, standard deviations and simple correlation coefficients) at the team level. Standard deviations, which now reflect only team (not individual) variance range from about a third of a scale point to a half of a scale point. Intercorrelations among dimensions at the team level are fairly high, and on average the shared variance among pair-wise comparisons of dimensions is 57 percent. This degree of covariance is not surprising, given that qualitative and inductive scale development procedures were not designed to yield orthogonal dimensions of effectiveness, and the instrument is designed to be a multidimensional measure of team effectiveness. Teams nevertheless differ from dimensions to dimension. For example, the team

that ranked the highest on goals ranked second on the customer dimension and third on the timeliness and quality dimensions. The team that ranked second on goals, ranked first on productivity, but seventh on the timeliness dimension. Thus, even in this small sample there is variance in terms of how teams perform on these various dimensions of effectiveness.

### Discussion

Research on work team effectiveness is challenged by the need to develop measures that can be used across organizations. The challenge is even greater for team research that additionally seeks to address teams in cross-cultural contexts. There has been little prior team research on measures that can be applied across organizations and less on instrumentation that can be applied cross-culturally. The study described in this paper sought to address these challenges by developing and then evaluating a measure to assess team effectiveness that can be completed by team members, leaders, and customers across cultures.

In doing so, we addressed two objectives in this research. The first of these was to develop a team effectiveness measure with wide generalizability. Structured interviews were conducted in six organizations in four countries. These interviews were designed to gain insight into team effectiveness items that are meaningful to a variety of teams across multiple contexts. Team effectiveness is likely to be multi-dimensional, therefore we sought to identify several dimensions of effectiveness based on these interviews. Dimensions and items were derived from interviews in a variety of cultural and organizational contexts, and selected to apply generally across teams performing different tasks, in different organizations and countries. A subset of items generated in this fashion formed internally consistent scales with reliability scores acceptable for new scales (Nunnally, 1978: 226). Once derived, the scales were subjected to a

rigorous translation-retranslation process in both Spanish and French. Analyses were then conducted to determine the suitability of the translations. Results suggest that scores are similar regardless of the language in which they are answered. Language version had only a small, non-statistically significant impact on the results.

Second, we sought to develop convergent instrument formats that could be utilized consistently across different types of raters. Convergence is an important indicator of construct validity (Campbell & Fiske, 1956; Schwab, 1980). The measures that emerged from our qualitative steps were further modified following reliability analyses that assessed internal consistency by rater groups. A second pilot test was performed on eleven teams with 99 members, leaders and customers to investigate convergence within scales and whether scale scores differed by team, which is desirable, and by role (member, leader or customer), which is undesirable. Results indicated significant variance between-teams. Furthermore, an individual's relationship to the team had no statistically significant impact on any of the scale scores beyond that explained by differences across teams. Given that the five scales measure dimensions of team effectiveness, positive intercorrelations between scales were expected. The dimensions were all positively correlated.

Overall, based on these analyses, the scales appear sensitive to variation in teams, and relatively insensitive to the source of the evaluation. Our final instrument appears in Appendix C. There are a variety of ways in which this instrument can be utilized for future research and practice. For example, given the high inter-correlation among the dimensions, for certain circumstances it might be appropriate to combine the five dimensions into a global measure of effectiveness. Such might be the case, for example, if a researcher is interested in predicting team effectiveness from a global set of independent variables. From a diagnostic point of view,

however, in which the objective is to help teams function more effectively across a variety of performance facets, analysis and feedback by dimension would be recommended.

### Limitations

The analyses conducted here can only be considered indirect assessments of construct validity since the criterion in any construct validity is conceptual (Schwab, 1980). Consideration should also be given to the definition of constructs and methods used in generating measures. We applied a rigorous method to define our construct, starting with a literature review to develop interview questions. We then conducted interviews in four cultures to derive items capturing general measures of team effectiveness. These methods strengthen the validity of our measures.

Even so, while accomplishing these objectives, we acknowledge that our study has several limitations. Our pilot samples were relatively small and prohibited the use of some statistical techniques such as structural equation modeling to test the invariance of the measurement model. We intend to confirm the model in our ongoing research with a large scale sample of teams in multinational organizations. Secondly, we tested only three language versions and thus cannot generalize beyond these languages. Future research should test the instruments in a broader sample of countries using additional language versions such as Chinese and Japanese. Finally, we acknowledge that although our samples include a diverse set of teams, we did not include an exhaustive variety of team types. It will be important in future research to confirm the applicability and validity of the measures in additional team contexts.

### Conclusion

In summary, this paper makes two substantial contributions to the literature. First, the measurement we have developed for teams can be utilized in future multinational management research investigating team effectiveness. Practitioners can deploy the assessment in order to

provide feedback and focus team effectiveness improvement efforts. Second, and perhaps more importantly, the process that we utilized to develop measures can be used by researchers interested in other constructs who are concerned about equivalence of meaning across cultures. The procedure employed here to develop measures of team outcomes cross-culturally appears promising. We were able to derive dimensions and items based on interviews. As we continue to refine these instruments, we hope that others will follow suit.

## References

- Ancona, D. G. (1990). Outward bound: Strategies for team survival in an organization. Academy of Management Journal, 33(2): 334-365.
- Ancona, D. G. & Caldwell, D. F. (1992a). Bridging the boundary: External activity and performance in organizational teams. Administrative Science Quarterly, 37: 634- 665.
- Ancona, D. G., & Caldwell, D. F. (1992b). Demography and design: Predictors of new product team performance. Organization Science, 3, 321-341.
- Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1967). Cross-cultural research methods. New York: Wiley.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relationships between work group characteristics and effectiveness: Implications for designing effective work groups. Personnel Psychology, 46, 823-850.
- Cascio, W. F. (1989). Managing human resources: Productivity, quality of work life, profits. New York: McGraw-Hill.
- Cohen, S. & Ledford, G. E. (1994). The effectiveness of self-managing teams: A quasi-experiment. Human Relations, 47(1), 13-43.
- Cohen, S.G., Ledford, G.E., & Spreitzer, G.M. (1996). A predictive model of self-managing work team effectiveness. Human Relations, 40, 643-675.
- Cohen, S.G., & Bailey, D.E. (1997). Trends in team effectiveness research: From the shop floor to the executive suite. Journal of Management, 23, 239-290.

Daft, R. L. (1997). Management. Fourth Edition. The Dryden Press.

Earley, P. C., & Erez, M. (1997). The Transplanted Executive. London: Oxford University Press.

Earley, P.C., & Gibson, C.B. (1998). Taking stock in our progress: 100 years of solidarity and community. Journal of Management, 2, 265-304.

Earley, P.C., & Singh, H. (1995). International and intercultural management research: What's next? Academy of Management Journal, 38, 327-340.

Gibson, C. B., & Kirkman, B. L. (1998). Our Past, present, and future in teams: The role of human resource professionals in managing team performance. In A.I. Kraut & A.K. Korman (Eds.), Changing concepts and practices for human resource management: Contributions from industrial-organizational psychology. San Francisco: Jossey-Bass.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. Administrative Science Quarterly, 29, 499-517.

Goodman, P. S., Devadas, R., & Griffith-Hughson, T. L. (1988). Groups and productivity: Analyzing the effectiveness of self-managing teams. In J.P. Campbell & R.J. Campbell (Eds.), Productivity in organizations: New perspectives from industrial and organizational psychology. San Francisco: Jossey-Bass.

Goodman, P. S., Ravlin, E. C., & Schminke, M. (1990). Understanding groups in organizations. In L.L. Cummings & B.M. Staw (Eds.), Leadership, participation & group behavior. (pp. 333-383). Greenwich, CT: JAI Press.

Hackman, J.R. (1986). Leading groups in organizations. In P.S. Goodman (ed.) Designing effective workgroups. San Francisco, CA: Jossey-Bass Publishers.

Hackman, J.R. (1990). Group influences on individuals in orgnaizations. In M.D.

Dunnette, and L.M. Hough, Handbook of Industrial Organizational Psychology Vol.3, 199-268  
to Alto, CA: Consulting Psychologist Press.

Hackman, J.R. & Oldham, G.R. (1975). Development of the Job Diagnostic Survey,  
Journal of Applied Psychology, 6(2): 159-170.

Hofstede, G. (1980). Culture's consequences: International differences in work related  
values. Beverly Hills: Sage Publications.

Jehn, K. A.(1995). A multimethod examination of the benefits and detriments of  
intragroup conflict. Administrative Science Quarterly. 40(2): 256 - 283.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a  
difference: A field study of diversity, conflict, and performance in workgroups. Administrative  
Science Quarterly, 44(4): 741-763.

Jewell, L., & Reitz, J. (1988). Group decision making. Ballinger Publishing Co./Harper &  
Row: Cambridge, MA  
Kabanoff, B. 1997. "Computers can read as well as count: Computer-  
aided text analysis in organizational research." Journal of Organizational Behavior, 18: 507-511.

Kirkman, B.L., & Shapiro, D.L. (1997). The impact of cultural values on employee  
resistance to teams: Toward a model of globalized self-managing work team effectiveness.  
Academy of Management Review, 22(3): 730-745.

Kirkman, B. L. & Rosen, B. (1999). Beyond self-management: Antecedents and  
consequences of team empowerment. Academy of Management Journal. 42(1): 58-74.

Kluckhohn, F. R., & Strodtbeck, F. L. (1961). Variation in value orientations. Evanston,  
IL: Row, Peterson & Company.

Lawler, III, E. E., Mohrman, S. A., & Ledford, Jr., G. E. (1995). Creating high  
performance organizations: Practices and results of employee involvement and total quality

management in fortune 1000 companies. San Francisco: Jossey Bass.

Maznevski, M. L., & DiStefano, J. J. (1997). Adding tools to the toolkit: Insights and empirical evidence on the value of additional cultural frameworks for international management research. Paper presented to the Academy of International Business, Monterrey, Mexico.

Mohrman, S. A., Cohen, S. G., & Mohrman, A. M., Jr. (1995). Designing Team-Based Organizations: New Forms for Knowledge Work. San Francisco: Jossey-Bass.

Nunnally, J. C. (1978). Psychometric Theory. McGraw-Hill Series in Psychology. New York: McGraw-Hill Book Company.

Osterman, P. (1994). How common is workplace transformation and who adopts it? Industrial and Labor Relations Review, 47, 173-188.

Saavedra, R., Earley, P. C., & Van Dyne, L. (1993). Complex interdependence in task performing groups. Journal of Applied Psychology, 78, 61-72.

Schwab, D. P. (1980). Construct validity in organizational behavior. Research in Organizational Behavior, 2, 3-43.

Shea, G. P., & Guzzo, R. A. (1987). Group effectiveness: What really matters? Sloan Management Review, Spring, 25-31.

Sundstrom, E., DeMeuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. American Psychologist, 45, 120-133.

Tsui, A. S. (1984). A role set analysis of managerial reputation. Organizational Behavior and Human Performance, 34, 64-96.

Van Maanen, J., & Barley, S. (1984). Occupational Communities, in B. Staw and L. L. Cummings (Eds.) Research in Organizational Behavior, vol. 6, (pp. 287-365). Greenwich, CT: JAI Press.

Wageman, R. (1995). Interdependence and group effectiveness. Administrative Science Quarterly, 40, 145-180.

Wall, T. D., Kemp, N. J., Jackson, P. R., & Clegg, C. W. (1986). Outcomes of autonomous workgroups: A longitudinal field experiment. Academy of Management Journal, 29, 280-304.

Watson, W. E., & Michaelson, L. K. (1988). Group interaction behaviors that affect performance on an intellectual task. Group and organization Studies, 13, 495-516.

Wellins, R. S., Byham, W. C., & Wilson, J. (1991). Empowered teams. San Francisco: Jossey-Bass.

Wellins, R. S., Byham, W. C., & Dixon, G. R. (1994). Inside teams : How 20 world-class organizations are winning through teamwork. San Francisco, CA: Jossey-Bass.

## Appendix A

### Journals Searched

- Journal of International Business Studies
- Columbia Journal of World Business
- Business Horizons
- Sloan Management Review
- California Management Review
- Harvard Business Review
- Fortune
- Business Week
- Academy of Management Journal
- Academy of Management Review
- Administrative Science Quarterly
- Journal of Applied Psychology
- Personnel Psychology
- Journal of Organizational Behavior
- Journal of Management
- Journal of Cross-Cultural Psychology
- Psychological Review
- Annual Review of Psychology
- Annual Review of Sociology
- Organizational Behavior and Human Decision Processes
- Psychological Bulletin

## Appendix B

### Interview Protocol

1. Could you tell us a little about what you do and the teams you work with?
2. Who is on the teams? How are these members selected? How are responsibilities divided?
3. What is the function of the teams [what outputs do they provide]?
4. Who is the team's "customer" [internal or external]?
5. Who receives the teams' work [who is directly downstream in the process]?
6. How is performance monitored and rewarded?
7. What kind of feedback do teams receive about performance?
8. How do you know when you have done a good job?
9. Do you believe the teams are effective? Why or why not?
10. Do the teams have leaders? What are the responsibilities of the leader?
11. Who does the team report to? Does it interact with other teams?
12. Would the teams benefit from more direction? Who should provide it? In what format?
13. What are the key factors that contribute to and/or inhibit the success of the teams?
14. How are practices shared in this organization?
15. To what extent does headquarters dictate practices?
16. Is individual achievement or collective achievement more important in this organization?
17. Is individual achievement or collective achievement more important in this country?
18. What facets of the culture here impact teams, either positively or negatively?
19. What metaphors [or mental images] do people use for teams in this country?
20. Do you have anything else you would like to add?

## Appendix C

## Team Outcome Effectiveness Items Included on Final Survey

---

1 = very inaccurate; 2 = mostly inaccurate; 3 = slightly inaccurate; 4 = uncertain; 5 = slightly accurate; 6 = mostly accurate; 7 = very accurate

GOALS

- |  |   |   |   |   |   |   |   |
|--|---|---|---|---|---|---|---|
| 1. This team fulfills its mission.                       | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. This team accomplishes its objectives.                | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. This team meets the requirements set for it.          | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. This team achieves its goals.                         | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. This team serves the purpose it is intended to serve. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

CUSTOMERS

- |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 1. This team's customers are satisfied.                         | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. This team's customers are happy with the team's performance. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. This team is responsive to its customers.                    | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. This team fulfills the needs of its customers.               | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. This team responds to external demands.                      | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

TIMELINESS

- |  |   |   |   |   |   |   |   |
|--|---|---|---|---|---|---|---|
| 1. This team meets its deadlines.  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. This team wastes time.  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. The team provides deliverables (e.g., products, or services) on time. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. This team is slow.  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. This team adheres to its schedule.                                    | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 6. This team finishes its work in a reasonable amount of time.           | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

QUALITY

- |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 1. This team has a low error rate.                      | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. This team does high quality work.                    | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. This team consistently provides high quality output. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. This team is consistently error free.                | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. This team needs to improve the quality of its work.  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
- 

Continued

Appendix C Continued

Team Outcome Effectiveness Items Included on Final Survey

---

1 = very inaccurate;2 = mostly inaccurate;3 = slightly inaccurate;4 = uncertain;5 = slightly accurate;6 = mostly accurate;7 = very accurate

PRODUCTIVITY

1. This team uses too many resources.	1	2	3	4	5	6	7
2. This team is productive.	1	2	3	4	5	6	7
3. This team is wasteful.	1	2	3	4	5	6	7
4. Inputs used by this team are appropriate for the outputs achieved.	1	2	3	4	5	6	7
5. This team is efficient.	1	2	3	4	5	6	7

---

Table 1

Initial Items Derived for From InterviewsItem

---

1. Put out product on time/Meet deadline/On-time delivery.
  2. Does the team answer/meet the needs of its customer(s)?
  3. Customer's level of satisfaction.
  4. Cycle time. Reduce cycle time.
  5. Has the team met the goals set for it? Achieving goals.
  6. Does the team have a common understanding of its purpose? Is the idea of the goal widely shared?
  7. Cost reduction.
  8. Error reduction/Lack of errors
  9. Improvement over prior performance on same problem/task. Continuous improvement/growth.
  10. Percentage of goal achieved.
  11. Implementation of decisions. Follow-through.
  12. Obtaining a sales target.
  13. Amount of re-work required/Error free work.
  14. Number of hours/days/etc. to complete task.
  15. Meeting/staying within budget.
  16. Does the team feel/believe it is successful?
  17. Successful at change. Did change happen that was planned/attempted? Take Action? Develop action items?
  18. Complaints/Lack of them/Reduction of them. From both internal and external customers.
  19. Productivity/Dollars spent against time used.
  20. Number of complaints/Customer complaints.
  21. Variation/Variance in Process or Performance.
  22. Quality.
  23. Market share compared to competitors.
  24. Shareholder satisfaction.
-

Table 2

Initial Dimensions

Dimension	Corresponding Items <sup>a</sup>
Meet Goals/Requirements:	(5) Has the team met the goals set for it? Achieving goals. (10) Percentage of goal achieved. (12) Obtaining a sales target. (15) Meeting/staying within budget.
Stakeholder Satisfaction:	(2) Does the team answer/meet the needs of its customer(s)? (3) Customer's level of satisfaction. (18) Complaints/Lack of them/Reduction of them. From both internal and external customers. (24) Shareholder satisfaction. (20) Number of complaints/Customer complaints.
Timeliness:	(1) Put out product on time/Meet deadline/On-time delivery. (4) Cycle time. Reduce cycle time. (14) Number of hours/days/etc. to complete task.
Errors:	(8) Error reduction/Lack of errors (13) Amount of re-work required/Error free work. (21) Variation/Variance in Process or Performance. (22) Quality.
Productivity:	(7) Cost reduction. (19) Productivity/Dollars spent against time used.

<sup>a</sup>Numbers correspond to item numbers in Table 1.

Continued

Table 2 Continued

Initial Dimensions

Dimension	Corresponding Items <sup>a</sup>
Change/ Implementation:	(9) Improvement over prior performance on same problem/task. Continuous improvement/growth.  (11) Implementation of decisions. Follow-through.  (17) Successful at change. Did change happen that was planned/attempted? Take Action? Develop action items?
Shared Understanding of Mission:	(6) Does the team have a common understanding of its purpose? Is the idea of the goal widely shared?
Market Share Compared to Competitor:	(23) Market share compared to competitors.

<sup>a</sup>Numbers correspond to item numbers in Table 1.

Table 3

Items Included on Pilot Survey


---

1 = very inaccurate; 2 = mostly inaccurate; 3 = slightly inaccurate; 4 = uncertain; 5 = slightly accurate; 6 = mostly accurate; 7 = very accurate

GOALS

1. This team fulfills its mission.	1	2	3	4	5	6	7
2. This team accomplishes its objectives.	1	2	3	4	5	6	7
3. This team meets the requirements set for it.	1	2	3	4	5	6	7
4. This team achieves its goals.	1	2	3	4	5	6	7
5. This team serves the purpose it is intended to serve.	1	2	3	4	5	6	7
6. This team adheres to its budget.	1	2	3	4	5	6	7

CUSTOMERS

1. This team's customers are satisfied.	1	2	3	4	5	6	7
2. This team's customers are happy with the team's performance.	1	2	3	4	5	6	7
3. This team rarely receives negative comments from its customers.	1	2	3	4	5	6	7
4. This team is responsive to its customers.	1	2	3	4	5	6	7
5. This team fulfills the needs of its customers.	1	2	3	4	5	6	7
6. This team responds to external demands.	1	2	3	4	5	6	7

TIMELINESS

1. This team meets its deadlines.	1	2	3	4	5	6	7
2. This team wastes time.	1	2	3	4	5	6	7
3. The team provides deliverables (e.g., products, or services) on time.	1	2	3	4	5	6	7
4. This team is slow.	1	2	3	4	5	6	7
5. This team adheres to its schedule.	1	2	3	4	5	6	7
6. This team takes a reasonable amount of time to complete its work.	1	2	3	4	5	6	7

---

Continued

Table 3 Continued

Items Included on Pilot Survey


---

1 = very inaccurate; 2 = mostly inaccurate; 3 = slightly inaccurate; 4 = uncertain; 5 = slightly accurate; 6 = mostly accurate; 7 = very accurate

QUALITY

1. This team makes mistakes.	1	2	3	4	5	6	7
2. This team has a low error rate.	1	2	3	4	5	6	7
3. This team does high quality work.	1	2	3	4	5	6	7
4. This team consistently provides high quality output.	1	2	3	4	5	6	7
5. This team is consistently error free.	1	2	3	4	5	6	7
6. This team needs to improve the quality of its work.	1	2	3	4	5	6	7

PRODUCTIVITY

1. This team uses too many resources.	1	2	3	4	5	6	7
2. This team is productive.	1	2	3	4	5	6	7
3. This team is wasteful.	1	2	3	4	5	6	7
4. This team rarely has cost over-runs.	1	2	3	4	5	6	7
5. Inputs used by this team are appropriate for the outputs achieved.	1	2	3	4	5	6	7
6. This team is efficient.	1	2	3	4	5	6	7

---

Table 4

Reliability and Item Analyses for a priori Scales

Scale	Customer Database <sup>a</sup>	Leader Database <sup>b</sup>	Member Database <sup>c</sup>	EF Database <sup>d</sup>	ES Database <sup>e</sup>	F Database <sup>f</sup>	S Database <sup>g</sup>
Goals	6	6	6	6	6	6	6
N	16	14	69	18	18	15	17
alpha	.72	.79	.89	.79	.75	.79	.63
Item(s) to drop to improve	6(.83)	6(.86)	6(.92)	6(.87)	6(.82)	6(.87)	6(.71)
alpha? (alpha if dropped)							
Customers	6	6	6	6	6	6	6
N	16	14	68	17	17	18	17
alpha	.94	.89	.91	.89	.79	.92	.70
Item(s) to drop to improve	3(.96)	3(.90)	[3(.90)]	[3(.83)]	3(.83)	[3(.914)]	3(.80)
alpha? (alpha if dropped)							

<sup>a</sup>Members = Multiple-rater pilot test: team members.

<sup>d</sup>EF = Bilingual Pilot Test: English Survey completed by French/English Bilinguals.

<sup>b</sup>Customers = Multiple-rater pilot test: team customers.

<sup>e</sup>ES = Bilingual Pilot Test: English Survey completed by Spanish/English Bilinguals.

<sup>c</sup>Leaders = Multiple-rater pilot test: team leader.

<sup>f</sup>S = Bilingual Pilot Test: Spanish Survey completed by Spanish/English Bilinguals.

<sup>g</sup>F = Bilingual Pilot Test: French Survey completed by French/English Bilinguals.

Continued

Table 4 Continued

Reliability and Item Analyses for a priori Scales

Scale	Customer Database <sup>a</sup>	Leader Database <sup>b</sup>	Member Database <sup>c</sup>	EF Database <sup>d</sup>	ES Database <sup>e</sup>	F Database <sup>f</sup>	S Database <sup>g</sup>
Timeliness	6	6	6	6	6	6	6
Number of items							
N	16	14	65	18	17	18	16
alpha	.85	.74	.85	.70	.77	.68	.76
Item(s) to drop to improve	2(.90)	4(.76)					
alpha? (alpha if dropped)							
Quality	6	6	6	6	6	6	6
Number of items							
N	16	14	68	18	18	18	16
alpha	.79	.78	.75	.89	.65	.91	.73
Item(s) to drop to improve	4(.79)	4(.79)	[1(.74)]	[1(.88)]	1(.72)	[1(.89)]	1(.80)
alpha? (alpha if dropped)	[1(.79)]	[1(.70)]					
Productivity	6	6	6	6	6	6	6
Number of items							
N	16	14	66	18	16	16	15
alpha	.76	.79	.80	.79	.57	.67	.60
Item(s) to drop to improve	[4(.76)]	1(.80)	[4(.78)]	4(.81)	4(.46)	4(.72)	4(.79)
alpha? (alpha if dropped)		[4(.80)]					

<sup>a</sup>Members = Multiple-rater pilot test: team members.

<sup>d</sup>EF = Bilingual Pilot Test: English Survey completed by French/English Bilinguals.

<sup>b</sup>Customers = Multiple-rater pilot test: team customers.

<sup>e</sup>ES = Bilingual Pilot Test: English Survey completed by Spanish/English Bilinguals.

<sup>c</sup>Leaders = Multiple-rater pilot test: team leader.

<sup>f</sup>S = Bilingual Pilot Test: Spanish Survey completed by Spanish/English Bilinguals.

<sup>g</sup>F = Bilingual Pilot Test: French Survey completed by French/English Bilinguals.

Table 5

Reliability for Final Scales

Scale	Customer Database	Leader Database	Member Database	EF Database	ES Database	F Database	S Database
Goals	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)
(Number of items in original scale)							
N	16	14	69	18	18	15	17
alpha	.83	.86	.91	.92	.85	.91	.73
Customers	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)
(Number of items in final scale)							
N	16	14	68	17	17	18	17
alpha	.96	.90	.90	.83	.85	.92	.81
Timeliness	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)
(Number of items in final scale)							
N	16	14	65	18	17	18	16
Alpha	.85	.74	.75	.88	.62	.88	.80

<sup>a</sup>Members = Multiple-rater pilot test: team members.

<sup>b</sup>Customers = Multiple-rater pilot test: team customers.

<sup>c</sup>Leaders = Multiple-rater pilot test: team leader.

<sup>d</sup>EF = Bilingual Pilot Test: English Survey completed by French/English Bilinguals.

<sup>e</sup>ES = Bilingual Pilot Test: English Survey completed by Spanish/English Bilinguals.

<sup>f</sup>S = Bilingual Pilot Test: Spanish Survey completed by Spanish/English Bilinguals.

<sup>g</sup>F = Bilingual Pilot Test: French Survey completed by French/English Bilinguals.

Continued

Table 5 Continued

Reliability for Final Scales

Scale	Customer Database	Leader Database	Member Database	EF Database	ES Database	F Database	S Database
Quality	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)
(Number of items in original scale)							
N	16	14	68	18	18	18	16
alpha	.76	.70	.75	.88	.62	.88	.80
Productivity	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)	6(5)
(Number of items in final scale)							
N	16	14	66	18	16	16	15
alpha	.76	.80	.78	.81	.46	.68	.79

<sup>a</sup>Members = Multiple-rater pilot test: team members.

<sup>b</sup>Customers = Multiple-rater pilot test: team customers.

<sup>c</sup>Leaders = Multiple-rater pilot test: team leader.

<sup>d</sup>EF = Bilingual Pilot Test: English Survey completed by French/English Bilinguals.

<sup>e</sup>ES = Bilingual Pilot Test: English Survey completed by Spanish/English Bilinguals.

<sup>f</sup>S = Bilingual Pilot Test: Spanish Survey completed by Spanish/English Bilinguals.

<sup>g</sup>F = Bilingual Pilot Test: French Survey completed by French/English Bilinguals.

Table 6

Bilingual Regression Results

Dependent Variable	Partial Regression Coefficients (t-statistics)				$\hat{R}^2$	(F)
	Spanish	French	Spanish	French		
Outcome Effectiveness						
Goals	.08 (.36)	-.00 (.00)	.00 (.00)	.00 (.07)	.00	(.07)
Customers	-.05 (-.22)	.06 (.26)	.06 (.26)	.00 (.08)	.00	(.08)
Timeliness	-.30 (-1.17)	-.04 (-.16)	-.04 (-.16)	.00 (.71)	.00	(.71)
Quality	-.07 (-.22)	.06 (.17)	.06 (.17)	.00 (.06)	.00	(.06)
Productivity	-.37 (-1.47)	.14 (.56)	.14 (.56)	.02 (1.69)	.02	(1.69)

Note: n=72.

\*p<.05.

\*\*p<.01.

Table 7

Regression Results on Teams and Roles

Dependent Variable	Teams Only		Partial Regression Coefficients (t-statistics)				Final		Change	
	$\hat{R}^2$	(F)	Customer	Leader	$\hat{R}^2$	(F)	$\hat{R}^2$	(F)		
Outcome Effectiveness										
Goals	.16	(2.82)**	.10	.27	(1.37)	.16	(2.51)**	.02	(1.04)	
Customers	.10	(2.10)*	-.27	.29	(1.27)	.12	(2.09)*	.03	(1.48)	
Timeliness	.16	(2.79)**	-.30	.14	(.49)	.15	(2.46)**	.01	(1.18)	
Quality	.22	(3.73)**	-.16	-.25	(-1.03)	.21	(3.19)**	.03	(1.86)	
Productivity	.19	(3.23)***	-.26	.31	(1.37)	.20	(3.08)***	.03	(1.60)	

Note: n=99.

\*p<.05.

\*\*p<.01.

\*\*\*p<.001.

Table 8

Descriptive Statistics for the Multiple Constituency Pilot

Variables	Mean	Std. Dev.	Simple Correlation Coefficients				
			Goals	Customer	Timeliness	Quality	Productivity
Goals	5.95	.34	1.00				
Customer	5.90	.35	.85	1.00			
Timeliness	5.30	.51	.65	.62	1.00		
Quality	5.14	.51	.81	.77	.53	1.00	
Productivity	5.67	.43	.89	.80	.79	.77	1.00

Note: n=11 teams. All zero-order correlation coefficients statistically significant,  $p < .01$ .