

Team Diagnostic Survey

Development of an Instrument

Ruth Wageman

Dartmouth College

J. Richard Hackman

Harvard University

Erin Lehman

Boston Adventures

This article describes the development of the Team Diagnostic Survey (TDS), an instrument intended for use both for the diagnosis of the strengths and weaknesses of work teams and for research on team behavior and performance. The TDS is based explicitly on existing research and theory about the conditions that foster team effectiveness. It provides an assessment of how well a team is structured, supported, and led as well as several indicators of members' work processes and their affective reactions to the team and its work. The psychometric properties of the TDS are satisfactory, based on analyses of data from 2,474 members of 321 teams in a diversity of organizations. Appropriate uses of the instrument are discussed.

Keywords: team diagnosis; team assessment; team design; team effectiveness

Researchers who seek to produce basic knowledge about organizational work teams that also can be used to help teams perform well face seemingly contradictory mea-

Support for the research reported here was provided by the Center for Public Leadership at Harvard University's Kennedy School of Government and by the MITRE Corporation. We gratefully acknowledge the assistance of numerous colleagues in obtaining access to the teams studied and to Michael O'Connor and the Hay Group for assistance in data collection.

THE JOURNAL OF APPLIED BEHAVIORAL SCIENCE, Vol. 41 No. 4, December 2005 373-398
DOI: 10.1177/0021886305281984
© 2005 NTL Institute

surement imperatives. The production of generalizable knowledge requires instruments that generate measures of team-relevant concepts that are both specific to research purposes and psychometrically sound (i.e., the measures are highly reliable and exhibit both convergent and discriminant validity). By contrast, instruments intended for use in helping teams perform better must focus on variables that are known to affect performance, can be manipulated, and are applicable across a variety of team types and settings so that norms can be developed for use in interpreting a team's scores.

Numerous instruments exist that meet either one or the other of these aspirations. There are, for example, literally dozens of consultant-developed instruments available for the diagnostic assessment of team dynamics (typical online examples include Cornelius Associates, 2004; Lefton & Buzzotta, 2005; Linkage Assessment Services, 2004; Reliable Surveys Online, 2005; and Parker's [1998] print compilation of 25 such tools). Typically, these instruments ask members to assess their teams on those dimensions that their developers assume to be most consequential for performance and most amenable to improvement through consultative intervention. Instruments of this type generally have high face validity and generate feedback and normative comparisons that teams and their leaders find interesting and informative. However, their content tends to be based more on the observations and inferences of practitioners than on established research and theory, and the factors they assess are not necessarily those that actually are most consequential for performance. Moreover, empirical findings obtained using such instruments are rarely reported in the research literature; as a consequence, they add little to basic knowledge about organizational work teams.

By contrast, scholar-developed measures of team attributes and dynamics abound in the research literature. Unlike consultant-developed instruments, such measures focus exclusively on those variables that are of research interest to the scholars who devise them; other factors that might be useful in generating a robust team diagnosis are not assessed. Because scholarly research requires that measures be highly reliable, scales tend to be lengthy and can take considerable time for respondents to complete. Moreover, scholars who study teams also desire measures that simultaneously are appropriate to the specific teams or organizations where data will be collected and that generate data comparable to that collected in related research. One common way of seeking to achieve both goals is to draw some items from existing measures but also to tailor or add items to make them more context specific. Such revisions unfortunately result in several different measures of essentially the same concept.

Ruth Wageman is an associate professor of business administration at the Amos Tuck School of Business, Dartmouth College. Her research focuses on identifying the organizational, group, and individual factors that influence the effectiveness of task-performing teams; the uses and misuses of power in teams; the antecedents and consequences of interdependence; and the theory and practice of leadership development.

J. Richard Hackman is Edgar Pierce Professor of Social and Organizational Psychology at Harvard University. His research interests include team dynamics, leadership effectiveness, and the design of self-managing social systems.

Erin Lehman is the chief operating officer of Boston Adventures, an organization specializing in corporate team building programs. She has developed executive education programs and conducted research for the Hauser Center for Nonprofit Organizations and the Center for Public Leadership at the John F. Kennedy School of Government and with the Careers Research Initiative at London Business School.

Measures of team task interdependence illustrate. Mohr (1971) and Pearce and Gregersen (1991) each generated measures of team task interdependence using a multi-item Likert-type scale. However, when these existing measures do not robustly capture the cooperation demands of a particular team task, researchers create revised scales that are more context specific (see, e.g., Campion, Medsker, & Higgs, 1993; Van der Vegt, Emans, & Van de Vliert, 2000, 2001). The net effect of such choices is that there is no standard measure of the concept—and therefore no possibility of developing norms that would aid in establishing the relative task interdependence of a given team. For all the aforementioned reasons, scholar-developed measures of team attributes generally are of limited use for team diagnosis or as guides for consultative interventions outside the context in which they are developed.

The Team Diagnostic Survey (TDS) is an instrument for assessing the properties of organizational work teams that has been specifically designed to be useful both in scholarly research on teams and in the practical diagnosis of teams' strengths and weaknesses. The TDS is explicitly based on a conceptual model of the factors that research has shown to be most consequential for team effectiveness. It is short enough to be completed by team members fluent in English in 20 minutes or less. Its contents were pretested using a wide variety of teams in a range of organizational settings, so it can be used to assess any type of work team that operates in an organizational context. It allows for the development of norms for different types of teams and organizations. It generates a diagnostic profile of a team that can help team members and leaders learn about the conditions that foster team effectiveness even as they explore the standing of their own team on those factors. And finally, the instrument's psychometric properties are strong enough for it to be used in scholarly research on team behavior and performance.

We describe in this article the conceptual model on which the TDS is based, provide a concrete description of the instrument, report its psychometric properties and preliminary data about its construct validity, and discuss its use both as a tool for the practical diagnosis of task-performing teams and in scholarly and evaluation research on team performance.

CONCEPTUAL BASIS OF THE INSTRUMENT

The TDS was designed to assess the concepts in the model of team effectiveness proposed by Hackman and his colleagues (Hackman, 1987, 1990, 2002; Hackman & Wageman, 2005). That model is unconventional in that it does not posit specific causal factors that are linked directly (or through mediating variables) to their intended effects. Neither does it identify factors that moderate the size or direction of specific input-output relations. Instead, work teams are viewed as semiautonomous social systems that evolve—over time, in their own idiosyncratic ways, and in interaction with their organizational contexts—into performing units that vary in effectiveness. As will be seen, the model identifies five conditions that when present increase the likelihood (but do not guarantee) that a team will perform well. The TDS provides measures of these enabling conditions. In addition, it assesses the effectiveness of team task pro-

cesses, the quality of members' work relationships, and individuals' motivation and satisfaction. We describe the theoretical basis for all these measures in the pages that follow.

Criteria of Team Effectiveness

Criterion measures in research on team performance often consist of whatever quantitative indicators as happen to be available or convenient to obtain. Such measures usually are unidimensional and overlook nonobvious indicators of how well a team is doing. We define team effectiveness more broadly, using the following three-dimensional conception:

1. The productive output of the team (that is, its product, service, or decision) meets or exceeds the standards of quantity, quality, and timeliness of the team's clients—the people who receive, review, and/or use the output. It is clients whose views count, not those of team members, except in those relatively rare cases when the team is the client of its own work.
2. The social processes the team uses in carrying out the work enhance members' capability to work together interdependently in the future. We define as effective only teams that are more capable as performing units when a piece of work is finished than they were when it was begun (Hackman, 1987, 1990, 2002; Hackman & Wageman, 2005).
3. The group experience contributes positively to the learning and well-being of individual team members rather than frustrating, alienating, or deskilling them.

The TDS generates direct measures of a team's standing on the second and third criteria but does not assess the first criterion because a self-report instrument cannot generate trustworthy data about the acceptability of a team's work to its clients. Therefore, the standing of teams on the first criterion is assessed indirectly, as we describe next.

Following Hackman (2002), we posit that a team's ability to generate products or services that are satisfactory to its clients is a joint function of the following three performance processes: (a) the level of effort group members collectively expend carrying out task work, (b) the appropriateness to the task of the performance strategies the group uses in its work,¹ and (c) the amount of knowledge and skill members bring to bear on the task. We refer to these three performance processes as process criteria of effectiveness (Hackman & Morris, 1975). Any team that expends sufficient effort in its work, deploys a task-appropriate performance strategy, and brings ample talent to bear on the work is quite likely to achieve a high standing on the first criterion of effectiveness specified earlier. By the same token, teams that operate in ways that leave one or more of these functions unfulfilled—that is, by expending effort insufficient for task demands, relying on performance strategies that are poorly aligned with task requirements, and/or bringing inappropriate or inadequate talent to bear on the work—are likely to fall short of client standards of acceptable performance. The TDS generates direct measures of a team's standing on each of the process criteria of effectiveness.

Enabling Conditions

The theory on which the TDS is based specifies that the chances for team effectiveness are higher when the following five conditions are met: (a) the people responsible

for the work are a real team rather than a team in name only, (b) the team has a compelling direction for its work, (c) the team's structure facilitates rather than impedes collective work, (d) the organizational context within which the team operates provides support for task activities, and (e) the team has available to it ample hands-on coaching to help members take the fullest possible advantage of their performance circumstances. These five conditions, each of which is made up of multiple subconditions, are described briefly next; for details, see Hackman (2002). Research attesting to the validity of the theory-specified conditions in influencing team effectiveness was reported by Hackman (1990), Hackman and O'Connor (2005), and Wageman (2001).

Real team. Real teams have three features. First, they have clear boundaries that reliably distinguish members from nonmembers. Second, team members are interdependent for some common purpose, producing a potentially assessable outcome for which members bear collective responsibility. Finally, real teams have at least moderate stability of membership, which gives members time and opportunity to learn how to work together well.

Compelling direction. The direction of a group is the specification of its overall purposes. Good team direction is challenging (which energizes members), clear (which orients them to their main purposes), and consequential (which engages the full range of their talents).

Also key to good direction for work teams is the emphasis direction setters give to the end-states to be achieved relative to the means that teams are to use in pursuing those ends. Specifically, task-focused motivation is fostered when ends are clearly and insistently specified but means are left to the team itself. Specifying both ends and means mitigates the challenge to the team and underemploys the full complement of team members' resources; specifying neither runs the risk of anarchy rather than focused, purposive collective work; and specifying means but not ends is likely to prompt the team to disengage from the work.

Enabling structure. Work teams often encounter difficulties that stem from not having been set up right. Sometimes the team's structure is overelaborate (which can create obstacles in getting things done); other times it is underspecified (a common problem for self-managing teams whose creators assume that teams should work everything out on their own); still other times, the problem is that the wrong structures are put in place. Our research has identified the following three structural features as key in fostering competent teamwork:

- Task design: The team task should be well aligned with the team's purpose and have a high standing on what Hackman and Oldham (1980) called "motivating potential." This means that the team task (a) is a whole and meaningful piece of work, (b) for which members have autonomy to exercise judgment about work procedures, and that (c) provides members regular and trustworthy knowledge of results. Well-designed group tasks foster high, task-focused effort by team members.
- Team composition: The team size should be as small as possible given the work to be accomplished, should include members with ample task and interpersonal skills, and should consist of a good diversity of membership—people who are neither so similar to one another that they duplicate one another's resources nor so different that they are unable to communicate or coordinate well with one

another. A well-composed team ensures that members have the full complement of knowledge and skills that is required to achieve the team's purposes and increases the ability of members to apply their complementary talents to the collective work.

- Core norms of conduct: A team should have established early in its life clear and explicit specification of the basic norms of conduct for member behavior. Shared expectations about behavior—whether “imported” by members from other group experiences, developed gradually as the team goes about its work, or explicitly specified by the person who creates the team—tend to remain in place until something fairly dramatic occurs to prompt members to reconsider what behaviors are and are not appropriate (Gersick, 1988). Team norms that foster good performance processes actively promote continuous scanning of the performance situation and proactive planning of group performance strategies, and they clearly specify those behaviors that are out of bounds for the team. Clear specification of core norms of conduct frees members from spending excessive time discussing the kinds of behavior that are acceptable in the group and facilitates the development of task performance strategies that are appropriate to the team's task and situation.

Supportive organizational context. Work teams sometimes find it difficult or impossible to obtain the kinds of organizational supports that are needed for effective performance, especially in established organizations whose human resource systems have been professionally designed and fine-tuned over the years to support work performed by individual employees. In addition to the material resources needed for actually carrying out the work, three features of the organizational context are especially consequential for work performed by teams.

First, the reward system should provide positive consequences for excellent team performance. Rewards and recognition to individuals should never be disincentives for task-oriented collaboration among team members, which is a common (if unintended) feature of traditional, individual-focused appraisal and compensation systems.

Second, the educational system should make available to the team, at the team's initiative, technical or educational assistance for any aspects of the work for which members are not already knowledgeable, skilled, or experienced—including if necessary the honing of members' skills in working together on collective tasks.

And third, the information system should provide the team with whatever data and projections members need to select or invent strategies for carrying out the work that are fully appropriate for the team's task and situation.

Available, expert coaching. Although coaching cannot compensate for a badly flawed team design, coaches can help teams take the best possible advantage of their performance circumstances (Hackman & Wageman, 2005; Wageman, 2001). The role of the coach is not to dictate to group members the one best way to proceed with the team's work. Instead, it is to help the team minimize its exposure to process losses (Steiner, 1972) and to maximize its chances of capturing synergistic process gains.

The specific kinds of help that coaches can provide for each of the three process criteria of effectiveness discussed earlier are as follows. For effort: helping members (a) minimize coordination and motivation problems and (b) build commitment to the group and its task. For performance strategy: helping members (a) avoid relying on habitual routines that may be inappropriate for the task or situation and (b) develop innovative ways of proceeding with the work that are well tuned to task requirements. For knowledge and skill: helping members (a) avoid inappropriate weighting of indi-

viduals' ideas and contributions and (b) share their expertise to build the team's repertory of skills.

The TDS assesses the degree to which a team's leader and members themselves provide the kinds of task-focused coaching just described. In addition, the instrument provides measures of the overall amount of coaching available to the team, members' perceptions of the helpfulness of that coaching, and for comparison, certain other leader behaviors not specified in the conceptual model, such as reinforcement of desired team behaviors, assistance with members' interpersonal relations, and activities that usurp the team's responsibility for managing its own work processes.

DEVELOPMENT OF THE TEAM DIAGNOSTIC SURVEY

The model just summarized is depicted in Figure 1. Our aspiration was to develop an instrument that reliably and efficiently assesses for any task-performing team the constructs in that model. As described next, we carried out two full revisions of a preliminary instrument using data collected from a large number of teams in a diversity of organizations, followed by a fine-tuning of the TDS based on psychometric analyses of the penultimate version, and finally by an initial assessment of the construct validity of the instrument.

Item Selection

The initial step in instrument development was to identify existing survey items and scales that assess the constructs of the conceptual model on which the TDS is based. Some items, especially those having to do with the design of the team's work and with member motivation and satisfaction, were drawn or adapted from research on the motivating properties of work designed for individuals (Hackman & Lawler, 1971; Hackman & Oldham, 1980). Other items, especially those having to do with the design and contextual features of teams, were drawn or adapted from an instrument designed for a cross-organizational study of team behavior and performance (Hackman, 1990) and from a precursor instrument developed by Wageman (2001) in a study of field service teams. Other items, especially those having to do with the description of leader behaviors, were adapted from research on leadership and group dynamics in professional symphony orchestras (Allmendinger, Hackman, & Lehman, 1996). And other items, especially those having to do with the quality of group interaction processes, were drawn from research on the structure and dynamics of airline flight deck crews (Hackman, 1993). We wrote new items for those constructs for which no satisfactory preexisting items were available.

The selection of items for multi-item scales always involves a trade-off between content diversity and internal consistency reliability (DeVellis, 1991). Selecting items that are highly similar to one another (e.g., those that ask essentially the same question in slightly different words) increases internal consistency reliability but lessens the substantive robustness of the composite score constructed by averaging them. Opting

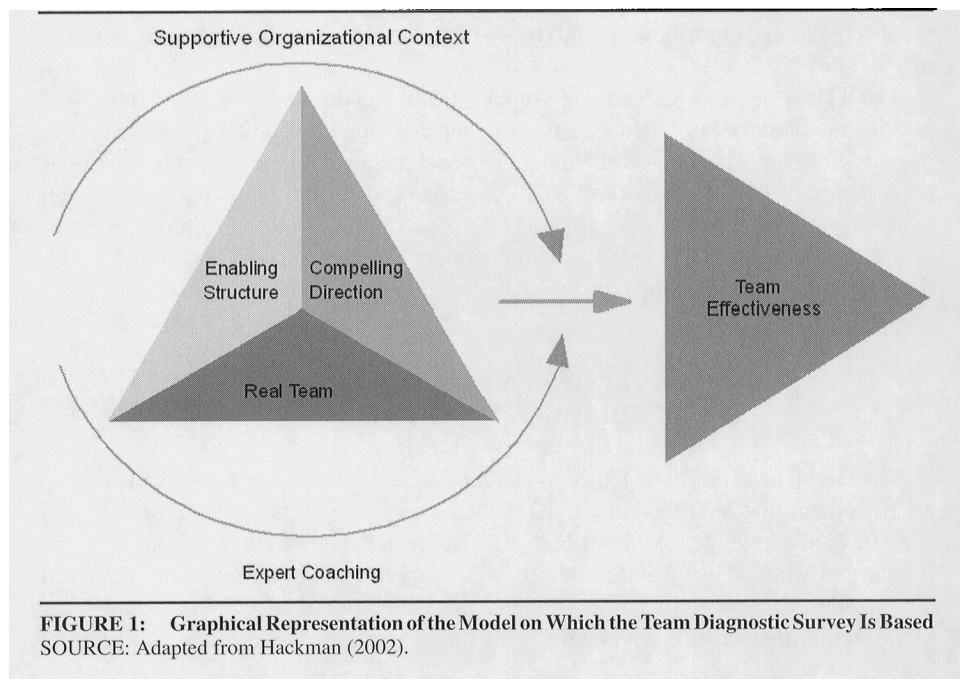


FIGURE 1: Graphical Representation of the Model on Which the Team Diagnostic Survey Is Based
 SOURCE: Adapted from Hackman (2002).

for diverse item content on the other hand increases the robustness of the composite measure but at a cost to reliability because interitem correlations will be lower. This trade-off can be mitigated by increasing the number of items used to construct the composite scores, but that strategy can result in an instrument so long that respondents find it burdensome to complete (Cortina, 1993; Ghiselli, 1964).

We sought to sample the domain of each construct widely but also to keep the TDS short enough to be completed in 15 to 20 minutes. This resulted in some scales that assess team-level properties having internal consistency reliabilities that would be marginal (e.g., .50 to .60) if the instrument were intended for assessing the views of individual respondents. In the present case however, all measures except those that assess individual motivation and satisfaction are intended for use only at the group level of analysis. Therefore, scale scores for variables that assess team attributes are constructed by averaging responses across both items and team members—in effect, treating each member as providing an additional estimate of the true score of the team on each scale. Following the recommendations of Office of Educational Assessment (1998) and Wuensch (2003), the reliability of the group-level composite scores was computed by applying the Spearman-Brown formula to the obtained reliabilities for individual respondents. To illustrate, consider a three-item scale administered to members of a four-person team. If the average intercorrelation among the three items were .25, internal consistency reliability at the individual level would be .50. But reliability

at the team level of analysis would be .80 (note that this number would be lower for teams smaller than four and higher for larger teams).

Analysis and Revision Process

The first draft of the TDS was administered to a convenience sample of 181 highly diverse teams from 63 different organizations (total number of respondents = 1,272). We calculated internal consistency reliabilities for all scales and flagged for revision those for which reliability at the individual level of analysis was less than .50.² To assess discriminant validity, we compared the average correlation of items within each scale with items from other scales and flagged for revision any scales for which the average within-scale correlation was less than .10 greater than the average inter-correlation among items assigned to different scales. Intraclass correlations were calculated for all scales to confirm that the scale scores could appropriately be interpreted at the group level of analysis.

Field researchers administering the first version of the TDS noted a few items that respondents found unclear or confusing, and these items were rewritten. Also, two substantial revisions were made in the section of the instrument that asks about the coaching received by teams. Namely, a section on the timing of leader coaching was dropped because it had poor psychometric properties, and a section that assesses peer coaching (as contrasted with coaching provided by team leaders) was added to ensure consistency with the construal of coaching specified in our model (Hackman & Wageman, 2005).

The revised version of the TDS was administered to a new sample of 140 teams, of which 76 were management teams and 64 were teams that carried out various kinds of analytic work (total number of respondents = 1,202). All scales in the revised version contained at least two items, and all composite scales had at least one item that was reverse-scored. We describe the contents of the revised instrument next and then report its psychometric properties.

DESCRIPTION OF THE TDS

The Team Diagnostic Survey is organized into 10 sections. Section 1 captures general descriptions of the team. Sections 2 through 7 assess the model-specified conditions for team effectiveness. Sections 8 and 9 provide measures of the three effectiveness criteria. Section 10 asks for respondent biographical information.

The model-specified scales assessed by the TDS are described next, including the specific items that comprise each construct.³ In the actual instrument, items are intermixed within sections. Except where otherwise noted, all items use a 5-point scale ranging from *highly inaccurate* (1) to *highly accurate* (5). Group-level composite scores are computed by averaging responses across items and respondents. Reverse-scored items are indicated by (R).

Real Team

A team's score for this construct is computed by averaging member responses on subdimensions that assess the degree to which the team is bounded, is stable over time, and has interdependent work.

Bounded

Team membership is quite clear—everybody knows exactly who is and isn't on this team.

There is so much ambiguity about who is on this team that it would be nearly impossible to generate an accurate membership list. (R)

Anyone who knows this team could accurately name all its members.

An additional indication of boundedness is obtained from member reports of team size. A smaller range (i.e., the difference between the largest and smallest reported size) is interpreted to reflect greater boundedness.

Stable

Different people are constantly joining and leaving this team. (R)

This team is quite stable, with few changes in membership.

Interdependent

Members of this team have their own individual jobs to do, with little need for them to work together. (R)

Generating the outcome or product of this team requires a great deal of communication and coordination among members.

Members of this team have to depend heavily on one another to get the team's work done.

Compelling Direction

A team's score is computed by averaging member responses on subdimensions that assess the degree to which team direction is clear, challenging, and consequential. An additional item assesses the focus of the team's direction on ends versus means.

Clear

There is great uncertainty and ambiguity about what this team is supposed to accomplish. (R)

This team's purposes are specified so clearly that all members should know exactly what the team exists to accomplish.

Challenging

This team's purposes are so challenging that members have to stretch to accomplish them.

This team's purposes are not especially challenging—achieving them is well within reach. (R)

Consequential

The purposes of this team don't make much of a difference to anybody else. (R)

This team's purposes are of great consequence for those we serve.

Ends Versus Means

Respondents check one of the following, and the team score is the modal response:

The purposes of our team are specified by others, but the means and procedures we use to accomplish them are left to us.

The means or procedures we are supposed to use in our work are specified in detail by others, but the purposes of our team are left unstated.

Both the purposes of our team and the means or procedures we are supposed to use in our work are specified in detail by others.

Neither the purposes nor the means are specified by others for our team.

Enabling Structure

A team's score is computed by averaging member responses on dimensions that assess the degree to which (a) the team is well composed, (b) the team's task is motivationally well designed, and (c) team norms of conduct are clear and well accepted.

Team composition. Subdimensions that are combined to assess the quality of a team's composition are (a) the degree to which the size of the team is appropriate for the work to be accomplished (i.e., neither too large nor too small), (b) the degree to which the team has an appropriate diversity of membership (i.e., members are neither too similar to one another nor too different from one another to accomplish team purposes), and (c) the level of member skills, including interpersonal skills.

Size

This team is larger than it needs to be. (R)

This team has too few members for what it has to accomplish. (R)

This team is just the right size to accomplish its purposes.

The composite score for team size is an algebraic combination of respondents' scores on the previous items, not a simple average as is computed for most other TDS scales. Specifically, the difference between the first two items is computed to assess the degree to which the team size is viewed as appropriate; this score can range from -4 (too small) to +4 (too large), with a score of 0 being optimum (neither too small nor too large). The absolute value of that difference score is then subtracted from 5 and averaged with the third item to produce a summary measure of team size that ranges from 1 to 5, with higher scores being more favorable.

Diversity

Members of this team are too dissimilar to work together well. (R)

This team does not have a broad enough range of experiences and perspectives to accomplish its purposes. (R)

This team has a nearly ideal "mix" of members—a diverse set of people who bring different perspectives and experiences to the work.

This composite score also is an algebraic combination of the items on which it is based, following the same logic as is used in scoring the group size items. Specifically, the difference between the first two items is computed to assess the degree to which the team has an appropriate level of diversity; this score can range from -4 (too homogeneous) to $+4$ (too heterogeneous), with a score of 0 being optimum. The absolute value of that difference score is then subtracted from 5 and averaged with the third item to produce a summary measure of team diversity that ranges from 1 to 5 , with higher scores being more favorable.

Skills

Members of this work team have more than enough talent and experience for the kind of work that we do.
Everyone in this team has the special skills that are needed for team work.
Some members of this team lack the knowledge and skills that they need to do their parts of the team's work. (R)

Team task design. Subdimensions that are combined to assess the design of a team's task are the degree to which the task is (a) a whole and meaningful piece of work that provides (b) autonomy for exercising judgment about work processes and (c) trustworthy knowledge of the results of the work.

Whole Task

We do a whole, identifiable piece of work.
Our team does such a small part of the overall task that it is hard to point specifically to *our* special contribution. (R)
This team's work is inherently meaningful.

Autonomy/Judgment

The work of this team leaves little room for the exercise of judgment or initiative. (R)
The work we do requires the team to make many "judgment calls" as we carry it out.

Knowledge of Results

Carrying out our team's task automatically generates trustworthy indicators of how well we are doing.
The work itself provides almost no trustworthy feedback about our team's performance. (R)
The only way we can figure out how well we are performing is for other people in the organization to tell us. (R)

Group norms. A team's score is computed by averaging member responses on the following items.

Group Norms

Standards for member behavior in this team are vague and unclear. (R)
It is clear what is—and what is not—acceptable member behavior in this team.
Members of this team agree about how members are expected to behave.

Supportive Organizational Context

A team's score is computed by averaging member responses on subdimensions that assess the availability of team rewards/recognition, information, training/technical

consultation, and basic material resources. Items are worded to provide descriptions of the organization in general, not just the proximal context of the focal team.

Rewards/Recognition

Excellent team performance pays off in this organization.

Even teams that do an especially good job are not recognized or rewarded by the organization. (R)

This organization recognizes and reinforces teams that perform well.

Information

It is easy for teams in this organization to get any data or forecasts that members need to do their work.

This organization keeps its teams in the dark about information that could affect their work plans. (R)

Teams in this organization can get whatever information they need to plan their work.

Education/Consultation

Teams in this organization have to make do with whatever expertise members already have—technical training and support are not available even when needed. (R)

When something comes up that team members do not know how to handle, it is easy for them to obtain the training or technical advice they need.

In this organization, teams do not receive adequate training for the work they have to do. (R)

Material Resources

Teams in this organization can readily obtain all the material resources that they need for their work.

Scarcity of resources is a real problem for teams in this organization. (R)

Available, Expert Coaching

The instrument addresses the following four aspects of team coaching: (a) the availability of coaching to the team, (b) the perceived helpfulness of the team leader's coaching, (c) the extent and focus of the team leader's coaching activities, and (d) the extent and focus of peer coaching among team members.

Availability of team coaching. Two measures of the availability of team-focused coaching are generated. The first is a single item that assesses the focus of the team leader's attention: Respondents rank the relative frequency with which their team leader engages in each of four kinds of activities, with the leader's relative emphasis on team coaching indicated by the mean rank of the second alternative.

Focus of Leader's Attention

Coaching individual team members.

Helping team members learn how to work well together.

Getting the team set up right—clarifying its purpose, picking members, structuring the task, setting expectations, and so on.

Running external interference for the team—getting resources, securing outside assistance, removing roadblocks, and so on.

The second measure of coaching availability is the mean of the following three items:

Coaching Availability

When members of teams in this organization have trouble working together, there is no one available to help them out. (R)

Teams in this organization have access to "coaches" who can help them learn from their successes and mistakes.

Expert coaches are readily available to teams in this organization.

Helpfulness of team leader coaching. Respondents are asked "Overall, how helpful is your team leader in building your team's capabilities?" The 5-category response scale ranges from *detrimental* through *neither particularly helpful or unhelpful* to *quite helpful*.

A summary indicator of helpful leader coaching is constructed from the types of helpful coaching the group receives and its availability. Computation is as follows. First, a composite score of leader helpfulness was computed by averaging the direct item listed earlier with measures described in the next subsection of the extent of leader task-focused coaching, leader operant coaching, and the reverse of leader unhelpful interventions (Wageman, 2001). The summary indicator is obtained by computing the square root of the product of the composite coaching availability and leader helpfulness scores. We chose this computation method because it makes the summary measure of helpful leader coaching comparable to the other TDS scales (i.e., it has a range from 1 to 5, with 5 being most favorable). Moreover, because the measure is based on the product of availability and helpfulness, a team cannot obtain a high score if either component is low, consistent with the theoretical model presented previously that specified that coaching must be both available and expert.

Extent and focus of team leader coaching. The TDS assesses the degree to which the team leader provides four different kinds of coaching: task-focused coaching, operant coaching (Komaki, 1998), interpersonal coaching, and unhelpful directives. Respondents are asked to indicate on a 4-point scale ranging from *never* to *often* how frequently the team leader engages in each behavior. For task-focused coaching, two items each assess coaching that focuses on team effort, performance strategies, and use of knowledge and skill, respectively. These three subscales are averaged to produce the overall measure of task-focused coaching.

Task-Focused Coaching

The team leader . . .

helps the team build a high shared commitment to its purposes.

helps the team sustain the motivation of all members.

works with the team to develop the best possible approach to its work.

keeps the team alert to anything that might require a change of work strategy.

helps members learn from one another and from the team's work experiences.

helps the team identify and use well each member's unique talents.

Operant Coaching

The team leader . . .

provides positive feedback when the team behaves or performs well.

provides corrective feedback when needed.

gives inappropriate or undeserved praise or criticism. (R)

Interpersonal Coaching

The team leader . . .

helps members resolve any conflicts that may develop among them.
helps members work on improving their interpersonal relationships.

Unhelpful Directives

The team leader . . .

micromanages the content and process of team discussions.
instructs the team in detail about how to solve its problems.
tells the team everything it is doing wrong.

Extent and focus of coaching provided by peers. Three types of peer coaching are assessed: task-focused coaching, interpersonal coaching, and unhelpful interventions (reverse-scored to produce an overall assessment of good-quality peer coaching). These items also use a 4-point scale ranging from *never* to *often*.

Task-Focused Peer Coaching

Regular team members . . .

take initiatives to promote high shared motivation and commitment.
take initiatives to make sure the team develops and uses the best possible approach to its work.
take initiatives to help the team build and use well members' knowledge and skills.

Interpersonal Peer Coaching

Regular team members . . .

take initiatives to constructively resolve any problems or conflicts that develop among members.

Unhelpful Peer Interventions

Regular team members . . .

tell other members what to do and how they should do it.

Process Criteria of Team Effectiveness

The instrument assesses the level of effort members collectively expend on the task, the quality of team task performance strategies, and the degree to which the team uses well the full complement of member knowledge and skill. The average of these three subscales provides an overall measure of the quality of team task processes.

Effort-Related Process Criteria

Members demonstrate their commitment to our team by putting in extra time and effort to help it succeed.
Everyone on this team is motivated to have the team succeed.
Some members of our team do not carry their fair share of the overall workload. (R)

Strategy-Related Process Criteria

Our team often comes up with innovative ways of proceeding with the work that turn out to be just what is needed.
Our team often falls into mindless routines, without noticing any changes that may have occurred in our situation. (R)
Our team has a great deal of difficulty actually carrying out the plans we make for how we will proceed with the task. (R)

Knowledge- and Skill-Related Process Criteria

How seriously a member's ideas are taken by others on our team often depends more on who the person is than on how much he or she actually knows. (R)

Members of our team actively share their special knowledge and expertise with one another.

Our team is quite skilled at capturing the lessons that can be learned from our work experiences.

Team Interpersonal Processes

The TDS includes two indicators of the quality of team members' interpersonal relations. The first measure assesses members' interactions; the second asks members about their satisfaction with within-team relationships.

Quality of Team Interaction

There is a lot of unpleasantness among members of this team. (R)

The longer we work together as a team, the less well we do. (R)

Working together energizes and uplifts members of our team.

Every time someone attempts to correct a team member whose behavior is not acceptable, things seem to get worse rather than better. (R)

Satisfaction With Team Relationships

My relations with other team members are strained. (R)

I very much enjoy talking and working with my teammates.

The chance to get to know my teammates is one of the best parts of working on this team.

Individual Learning and Well-Being

The instrument includes the following three measures of individual affective reactions to the team and its work: level of internal work motivation, satisfaction with growth opportunities, and overall satisfaction with the team. Each of these items and the items assessing satisfaction with team relationships (previous list) are measured on a 5-point scale ranging from *strongly disagree* to *strongly agree*.

Internal Work Motivation

I feel a real sense of personal satisfaction when our team does well.

I feel bad and unhappy when our team has performed poorly.

My own feelings are not affected one way or the other by how well our team performs. (R)

When our team has done well, I have done well.

Satisfaction With Growth Opportunities

I learn a great deal from my work on this team.

My own creativity and initiative are suppressed by this team. (R)

Working on this team stretches my personal knowledge and skills.

General Satisfaction

I enjoy the kind of work we do in this team.

Working on this team is an exercise in frustration. (R)

Generally speaking, I am very satisfied with this team.

Supplementary Measures

The TDS includes a small number of additional measures that are not included in our conceptual model of team effectiveness but that can be helpful in interpreting a team's standing on the model-specified constructs.

1. Descriptive information about the team being assessed, including the purpose of the team, the number of men and women in the team, the expected life span of the team, whether the work of the team is the main part of the respondent's job or is only a small part of the job, and whether the respondent is the formal leader of the team.
2. The level of authority the team has to manage its work. Respondents circle yes or no to indicate whether or not the team has the authority to decide about specific work-related matters. These responses are scored to place the team in one of the following five categories:
 - Self-governing team: The team has the authority to manage its work processes, alter its design and organizational features, and specify its main purposes.
 - Professional team: The team has the authority to specify its main purposes but not to alter its design or organizational features.
 - Self-designing team: The team has the authority to alter design or organizational features that are affecting the team or its work (e.g., selecting its own members, obtaining special training, etc.).
 - Self-managing team: The team has the authority to manage its own work processes and to change or adjust them if needed.
 - Manager-led team: The team has no authority beyond actually carrying out the work.
 - Hybrid team: The team does not fall into any of these categories.
3. The final optional section of the TDS requests brief biographical information from respondents. Questions ask for the respondent's age, gender, tenure in the organization, time in present job, and time as a member of the team being described.

EMPIRICAL PROPERTIES OF THE TEAM DIAGNOSTIC SURVEY

Psychometric information on the TDS scales for the 140 teams whose members completed the second revision of the instrument are presented for the five enabling conditions, the coaching variables, and the criterion variables in Tables 1 through 3. In addition to the mean and standard deviation of each variable, we report the average correlation of items within each scale (labeled *r within scale*) and the average correlation of items with all other items in the same section of the TDS (labeled *r between scale*). Comparison of the within and the between correlations provides a conservative test of discriminant validity because items within sections are most likely to be subject to method variance and because TDS items within sections assess constructs that are closely related conceptually. We also report both internal consistency reliabilities (coefficient alphas) for data from individual respondents (even though the instrument is not intended for use in assessing individual perceptions) and the projected reliabilities of team scores (conservatively assuming teams of size four).

As Table 1 shows, all team-level reliabilities are satisfactory, although they are relatively low for some subcomponents of two structure variables: size and diversity for team composition and whole task and autonomy/judgment for team task design.

TABLE 1
Descriptive Statistics and Psychometrics for the Five Enabling Conditions

	<i>Items</i>	<i>M</i>	<i>SD</i>	<i>r Within Scale</i>	<i>r Between Scale</i>	<i>Alpha (Individual)</i>	<i>Alpha (Team)</i>
Real team		3.94	0.66				
Bounded	2	4.41	0.85	.40	.15	.57	.84
Interdependent	2	3.90	0.91	.39	.11	.56	.84
Stable	2	3.52	1.18	.68	.12	.81	.94
Compelling direction		3.86	0.61				
Clarity	2	3.64	1.09	.46	.05	.63	.87
Challenge	2	3.57	0.98	.21	-.07	.35	.68
Consequentiality	2	4.38	0.84	.51	.15	.68	.89
Enabling structure		3.64	0.59				
Team composition	7	3.71	0.68				
Size	2	3.51	1.18	.22	.10	.36	.69
Diversity	2	3.88	0.79	.18	.18	.31	.64
Skills	3	3.74	0.87	.36	.20	.63	.87
Task design	7	3.81	0.63				
Whole task	2	3.75	0.93	.19	.18	.32	.65
Autonomy/judgment	2	4.09	0.77	.25	.09	.40	.73
Knowledge of results	3	3.59	0.87	.41	.19	.68	.89
Group norms	3	3.41	1.01	.55	.21	.79	.94
Supportive context		3.33	0.73				
Rewards/recognition	3	3.56	0.97	.58	.31	.81	.94
Information	3	3.32	0.90	.48	.36	.73	.92
Education/consultation	3	3.45	0.88	.48	.35	.73	.92
Material resources	2	2.98	0.98	.48	.33	.65	.88
Team coaching		3.29	0.69				
Coaching availability	3	3.03	0.94	.50	.31	.75	.92
Coaching helpfulness	1	3.66	0.60				

NOTE: *N* = 1,202 respondents in 140 teams. Projected internal consistency reliabilities (coefficient alphas) for the team level of analysis are based on an assumption of a team of four members; reliabilities would be larger for larger teams and smaller for smaller teams.

Discriminant validities are unsatisfactory for the diversity and whole task variables, which suggests that the subcomponents of team composition and task design should be interpreted with caution.

For conceptual reasons, we did not calculate internal consistency reliabilities or average intercorrelations for the overall real team, direction, enabling structure, and supportive context constructs. In each case, the elements of these higher level constructs can (and do) vary independently of each other. For example, a team's direction could be quite clear but neither challenging nor consequential; alternatively, it could be highly consequential but not at all clear or challenging. In such cases, the fact that a team scored higher on some components of the higher level construct than others would be an empirical finding of diagnostic significance for the focal team rather than a reliability problem. The same logic applies to the other global conditions whose components can vary independently.

TABLE 2
Descriptive Statistics and Psychometrics for Leader and Peer Coaching Activities

	<i>Items</i>	<i>M</i>	<i>SD</i>	<i>r Within Scale</i>	<i>r Between Scale</i>	<i>Alpha (Individual)</i>	<i>Alpha (Team)</i>
Leader coaching							
Task-focused coaching	6	3.67	.76	.66	.41	.92	.98
Operant coaching	2	3.63	.79	.27	.38	.43	.75
Interpersonal coaching	2	3.12	.92	.65	.47	.79	.94
Unhelpful directives	3	3.27	.74	.34	.33	.51	.86
Peer coaching							
Task-focused coaching	3	3.40	.78	.76	.38	.90	.97
Interpersonal coaching	1	3.16	.94				
Unhelpful interventions	1	2.65	.91				

TABLE 3
Descriptive Statistics and Psychometrics for Team Effectiveness Criteria

	<i>Items</i>	<i>M</i>	<i>SD</i>	<i>r Within Scale</i>	<i>r Between Scale</i>	<i>Alpha (Individual)</i>	<i>Alpha (Team)</i>
Process criteria							
Effort	3	3.80	.89	.49	.41	.74	.92
Performance strategy	3	3.58	.85	.43	.42	.69	.90
Knowledge and skill	3	3.52	.81	.40	.40	.67	.89
Team social processes							
Quality of interaction	4	3.80	.79	.47	.41	.78	.93
Relationship satisfaction	3	3.91	.70	.52	.33	.76	—
Individual well-being							
Internal work motivation	4	4.13	.60	.33	.23	.66	—
Growth satisfaction	3	3.87	.77	.45	.37	.71	—
General satisfaction	4	3.89	.84	.57	.38	.84	—

NOTE: Team-level reliabilities were not computed for the measures of motivation and satisfaction because these are individual-level constructs.

Table 2 presents psychometric data for the coaching measures. The two-item measure of leader operant coaching and the three-item measure of leader unhelpful directives exhibit weak discriminant validities and have internal consistency reliabilities that are lower than those of the other coaching measures. Table 3 presents psychometric data for the criterion measures. All reliabilities are satisfactory, but discriminant validities are weak both for the measures of the three process criteria and for the measure of the quality of member interaction. It may be that this reflects a substantive reality—namely, that teams that are doing well tend to have good processes across the board (Hackman, 1987). Alternatively, it may be that the instrument is not sensitive enough to distinguish cleanly among the several aspects of group process. The present data do not allow us to choose between these two explanations. Only alter-

native and independent assessments of group processes (e.g., by coded observations) would ultimately allow one to distinguish between the two possibilities.

We retained the coaching and process criteria subscales because of their importance in the model and because similar versions of these scales have in other studies achieved sufficient discriminant validity and have been important predictors of team effectiveness (see, e.g., Wageman, 2001; Wageman & Mannix, 2005).

Intraclass correlations were calculated for all measures to ensure that it is appropriate to aggregate and interpret them at the group level of analysis. The intraclass correlations ranged from a high of .63 (for compelling direction) to a low of .24 (for respondent gender, an inherently individual-level variable), with a median of .50. All were significant at $p < .001$. In addition, correlations among the five enabling conditions were examined. They ranged from .19 (real team with coaching availability) to .59 (context supports with coaching availability), with a median of .38, indicating that the enabling conditions are positively but moderately associated with one another, as would be expected.

Based on the data reported earlier and on comments from colleagues who had used the penultimate version of the TDS, we made several minor revisions of items.⁴ The final version of the instrument, presented in the previous section of this article, provides reasonably sound measures of the constructs in the model of team effectiveness on which the TDS is based.

DISCUSSION

The Team Diagnostic Survey has adequate psychometric properties, is useable for any type of task-performing team, and can be completed in 20 minutes or less by team members who are fluent in English. There remain, however, questions about the substantive validity of the TDS and about its appropriate use that can be resolved only as researchers and practitioners accumulate additional experience with the instrument.

Validity of the TDS

Although the previously reported intraclass correlations for constructs assessed by the TDS document that the instrument reliably distinguishes among teams, it also is necessary to demonstrate that it can detect meaningful differences among different types of teams and that it can predict how well teams actually perform. We discuss these matters next.

Findings about the power of the TDS to detect substantively meaningful differences across team types are provided in Table 4. The sample consisted of three types of teams: coacting intelligence analysis groups whose members work in parallel, interdependent analytic teams whose members work together to generate intelligence products for which the team as a whole is accountable, and senior executive teams whose members share responsibility for overseeing the performance of large organizations. As is seen in the table, statistically reliable differences across team types were obtained for 28 of the 36 TDS composite scores. The interdependent teams score

TABLE 4
Comparison of Composite Score Means for Three Types of Work Teams

	<i>Coacting Analytic (n = 38)</i>		<i>Interdependent Analytic (n = 26)</i>		<i>Senior Management</i>		<i>F Value</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Real team	3.61	.37	4.10	.37	4.11	.34	26.22***
Bounded	4.30	.35	4.50	.53	4.51	.41	3.18*
Interdependent	3.49	.61	4.09	.48	4.02	.38	18.55***
Stable	3.05	.65	3.71	.75	3.80	.63	17.03***
Compelling direction	3.59	.29	3.84	.34	3.99	.37	17.48***
Clarity	3.54	.62	3.99	.72	3.64	.62	4.14*
Challenge	3.13	.52	3.09	.61	3.83	.45	35.18***
Consequentiality	4.10	.49	4.45	.37	4.50	.38	12.75***
Enabling structure	3.55	.35	3.78	.33	3.67	.31	3.95*
Team composition	3.46	.40	3.70	.44	3.83	.32	13.28***
Task design	3.74	.36	4.00	.39	3.82	.33	4.22*
Group norms	3.46	.56	3.65	.49	3.36	.52	2.83
Supportive context	3.21	.46	3.32	.45	3.41	.36	3.19*
Rewards/recognition	3.28	.67	3.53	.58	3.73	.42	9.20***
Information	3.17	.46	3.25	.54	3.43	.44	4.35**
Education	3.43	.57	3.43	.42	3.48	.43	0.21
Resources	2.96	.50	3.08	.67	3.02	.51	0.35
Team coaching	3.20	.37	3.40	.37	3.30	.34	2.50
Coaching availability	2.94	.39	3.08	.51	3.06	.45	1.12
Coaching helpfulness	3.61	.40	3.82	.37	3.65	.30	3.23*
Leader coaching activities							
Task focused	3.58	.52	3.91	.51	3.65	.35	4.95**
Operant	3.62	.54	3.77	.68	3.58	.38	1.48
Interpersonal	3.11	.62	3.40	.58	3.08	.41	4.00*
Unhelpful	3.26	.52	3.35	.62	3.24	.36	0.57
Peer coaching activities							
Task focused	3.44	.47	3.88	.46	3.30	.34	19.79***
Interpersonal	3.13	.48	3.50	.74	3.16	.39	6.25**
Unhelpful	2.56	.53	2.65	.50	2.66	.35	0.69
Process criteria	3.69	.51	4.03	.48	3.58	.35	10.80***
Effort	3.65	.64	4.05	.58	3.87	.42	4.84**
Strategy	3.73	.51	4.10	.52	3.44	.39	21.84***
Knowledge and skill	3.69	.50	3.93	.46	3.43	.36	14.80***
Team social processes							
Quality of interaction	3.76	.51	4.06	.49	3.79	.39	4.21*
Relationship satisfaction	3.85	.44	4.16	.40	3.87	.30	7.15***
Individual well-being							
Internal work motivation	3.87	.29	4.14	.33	4.23	.27	20.08***
Growth satisfaction	3.88	.48	4.14	.43	3.84	.33	5.74**
General satisfaction	3.76	.53	4.18	.63	3.91	.41	5.82**

NOTE: $df = 2,137$.* $p < .05$. ** $p < .01$. *** $p < .001$

higher on the enabling conditions and on the criterion measures than do the coaching groups. The senior management teams score especially high on some aspects of team direction and contextual supports and on member internal motivation but are lower on clarity of group norms and on some aspects of coaching—especially task-focused peer coaching. These data of course are merely illustrative; only as additional TDS data accumulate for a broader sample of teams and organizations will it become possible to establish stable norms for teams of different types.

Additional data also will be required to fully establish the construct validity of the instrument by documenting the antecedents and consequences of the model-specified concepts assessed by the TDS as well as other aspects of group functioning with which a team's standing on the TDS may be associated. Two existing studies do provide supportive evidence regarding the construct validity of the instrument. Wageman (2001) used many scales from a precursor of the TDS in a study of field service teams at the Xerox Corporation and found that the scales significantly predicted both level of team self-management and objectively measured team performance. For these teams, direction, structure, and context together—which she termed *design features* of the teams—controlled significantly more criterion variance than did team leaders' coaching behaviors (for team self-management, design features controlled 42% of the variance, compared to less than 10% for the coaching measures; for objective performance, design controlled 37% of the variance, compared to less than 1% for coaching).⁵

In a recent study that used the penultimate version of the TDS to assess the intelligence analysis teams mentioned earlier, Hackman and O'Connor (2005) found that TDS-generated measures of the enabling conditions significantly predicted team effectiveness: In a linear regression, the five conditions controlled 74% of the variation of a criterion measure constructed by averaging three different effectiveness indicators. In this study, coaching activities were much stronger predictors of effectiveness than was the case in the Wageman (2001) study; indeed, task-focused coaching by peers had the highest correlation with the composite effectiveness criterion of any of the TDS measures.

The findings summarized here show that validation of the TDS and refinement of the theory on which it is based must proceed simultaneously as research continues to identify which of the enabling conditions are most potent for which types of teams and why different kinds of teams are helped more by certain conditions than by others. Moreover, it will be informative to explore in detail the dynamics of teams for which the TDS-measured enabling conditions do not predict effectiveness—that is, teams that do well despite their poor performance circumstances and teams that perform poorly despite the presence of the enabling conditions.

Using the TDS

The Team Diagnostic Survey was designed for two mutually supportive activities: research on the effectiveness of task-performing teams and interventions in social systems that seek to improve team effectiveness. A paper version of the instrument and the key for computing group-level composite scores is available from the authors, and

an online version of the instrument suitable for both research and consultative applications is available at <http://research.wjh.harvard.edu/TDS>.

Data from the TDS—whether collected by researchers, consultants, team leaders, or members themselves—provide reliable indicators of the quality of team design and leadership and can identify the major points of leverage (e.g., an overlarge team or lack of coaching) for improving a team's effectiveness. For example, Table 4 shows that a number of the senior management teams in our sample suffer from unformed norms of conduct. The CEOs of such teams might be coached in how to help the team articulate and enforce constructive operating norms.

Comparisons across teams in an organization can be used to pinpoint commonalities (e.g., poor direction setting or few rewards for team excellence) that could focus training and development activities for team leaders or organizational system redesign. For example, Table 4 shows that the coaching analytic groups are underchallenged by their purposes and are underrewarded as teams. An organizational change intervention aimed at helping leaders create real teams, teaching them to articulate challenging purposes, and redesigning the reward system to provide valued outcomes for team excellence could have a major positive impact on team effectiveness in this organization.

Note that the Team Diagnostic Survey is intended for use only with intact teams that perform work in a social system context, not groups that are teams in name only. Because TDS scores are reliable and meaningful only at the team level of analysis, the instrument is not appropriate for assessing the perceptions or attitudes of individual members. Data from team members should always be aggregated to the team level, and intraclass correlations should always be computed to affirm that the team-level scores can be meaningfully interpreted.

When used to diagnose the strengths and weaknesses of work teams, the aspiration should be to obtain responses from all team members because those who prefer not to complete the instrument may have perceptions that would be critical in generating a robust assessment of the team. We suggest that data for a team not even be analyzed unless more than 50% of its members—and ideally substantially more than that—complete the survey. The instrument also is not appropriate for personnel evaluation, such as to assess how well team leaders are doing their jobs. Instead, the greatest practical benefits of using the TDS are obtained when team members meet to review their TDS findings, explore the possible reasons for the team's pattern of scores, and reflect together about what they might do to strengthen their team's standing on the enabling conditions. The online version of the TDS, cited earlier, provides participating teams with a set of feedback charts and discussion questions specifically intended to help members have constructive discussions about such matters.

When the TDS is used in scholarly or evaluation research, independent assessments should always be obtained for those aspects of team behavior or performance that are of greatest research interest, as was done in the Hackman and O'Connor (2005) study described earlier. Independent assessments are important in part because of the vulnerability of self-report surveys to interpretive ambiguities that stem from common method variance. But it also is the case that many features of teams—especially team processes, leadership, and performance—are vulnerable to cognitive and affective

distortions of which respondents may be unaware. Proper interpretation of TDS findings require that aggregated individual responses be anchored in objective reality.

Moreover, because teams are social systems, they often create and enforce social realities that may be at variance with the actual features of their performance circumstances. Although social realities may be important to understand for some research purposes, it also is critical to know how well respondents' views fit with indicators that are less vulnerable than surveys to the influence of lay theories about team dynamics and performance (Staw, 1975).

Whether used for team diagnosis or scholarly research, the Team Diagnostic Survey can bring quantitative data to bear in enriching understanding of work teams and in correcting the distortions that pervade the everyday assessments made both by team members and by outside observers. Even so, findings from the TDS itself can be misleading if care is not taken to ensure that the conditions of data collection, data analysis procedures, and interpretations of TDS findings all are managed at the highest levels of professionalism.

NOTES

1. A team's performance strategy is the set of choices members make about how to carry out the work. For example, a manufacturing team might decide to divide itself into three subgroups, each of which would produce one subassembly, with the final product to be assembled later. Or a basketball team might decide to use modified zone defense, with one player assigned to guard the opposing team's best shooter. Or a team performing a creative task might choose to free associate about possible solutions in the first meeting, reflect for a week about the ideas that came up, and then reconvene to draft the product. All of these are choices about strategy.

2. Interitem correlations on which reliabilities were based were computed both (a) using the entire sample of 1,272 respondents and (b) using multiple subsamples, each of which consisted of one respondent from each of the 181 teams (the latter procedure is not vulnerable to the nonindependence of the responses of multiple individuals within given teams). Because these two procedures yielded virtually identical interitem correlations, we report the simpler analysis that uses all 1,272 respondents.

3. The items presented are those included in the final version of the Team Diagnostic Survey after fine-tuning a few items based on analyses of the second revision of the instrument.

4. Specifically, we revised one item in the challenge scale because comments from users suggested that respondents were interpreting the item as a statement about team skills rather than about team purposes. We added one item each to the scales assessing team boundedness and interdependence. We revised one item each from the size and judgment scales, revised both items in the diversity scale to better reflect the construct definition, and added one item to the whole task scale. We also revised one item each from the leader operant coaching and leader unhelpful directives scales to address their reliability problems. Finally, we deleted one item from the general satisfaction scale.

5. Moreover, competent task-focused coaching (e.g., leading a problem-solving process) helped well-designed teams exploit their favorable circumstances but made little difference for poorly designed teams. By contrast, unhelpful coaching interventions (e.g., identifying a team's problems and telling members how they should solve them) were much more deleterious for poorly designed teams than for those that had an enabling team structure and a supportive organizational context.

REFERENCES

- Allmendinger, J., Hackman, J. R., & Lehman, E. V. (1996). Life and work in symphony orchestras. *Musical Quarterly*, 80, 194-219.
- Campion, M. A., Medsker, G. J., & Higgs, C. (1993). Relations between work group characteristics and efficiency: Implications for designing effective work groups. *Personnel Psychology*, 46, 823-850.
- Cornelius Associates. (2004). *Team roles analysis: Self-assessment form for team building*. Available at <http://www.corneliusassociates.com>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, 31, 9-41.
- Ghiselli, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.
- Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice Hall.
- Hackman, J. R. (1990). *Groups that work (and those that don't)*. San Francisco: Jossey-Bass.
- Hackman, J. R. (1993). Teams, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 47-69). San Diego, CA: Academic Press.
- Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston: Harvard Business School Press.
- Hackman, J. R., & Lawler, E. E. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology Monograph*, 55, 259-286.
- Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 45-99). San Diego, CA: Academic Press.
- Hackman, J. R., & O'Connor, M. (2005). *What makes for a great analytic team? Individual vs. team approaches to intelligence analysis*. Washington, DC: Intelligence Science Board, Office of the Director of Central Intelligence.
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.
- Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, 30, 269-287.
- Komaki, J. L. (1998). *Leadership from an operant perspective*. New York: Routledge.
- Lefton, R. E., & Buzzotta, V. R. (2005). *The Team Scorecard*. Available at <http://www.communicare.com>
- Linkage Assessment Services. (2004). *The HiPlay Team Assessment*. Burlington, MA: Linkage, Inc.
- Mohr, L. B. (1971). Organizational technology and organizational structure. *Administrative Science Quarterly*, 16, 444-459.
- Office of Educational Assessment. (1998). *Inter-rater reliability*. Seattle: University of Washington.
- Parker, G. R. (1998). *25 instruments for team building*. Amherst, MA: HRD Press.
- Pearce, J. L., & Gregersen, H. B. (1991). Task interdependence and extra role behavior: A test of the mediating effects of felt responsibility. *Journal of Applied Psychology*, 76, 838-844.
- Reliable Surveys Online. (2005). *Team effectiveness*. Available at <http://reliablesurveys.com>
- Staw, B. M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior and Human Performance*, 13, 414-432.
- Steiner, I. D. (1972). *Group process and productivity*. San Diego, CA: Academic Press.
- Van der Veegt, G. S., Emans, B. J., & Van de Vliert, E. (2000). Affective responses to intragroup interdependence and job complexity. *Journal of Management*, 26, 633-655.
- Van der Veegt, G. S., Emans, B. J., & Van de Vliert, E. (2001). Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. *Personnel Psychology*, 54, 51-68.

- Wageman, R. (2001). How leaders foster self-managing team effectiveness: Design choices versus hands-on coaching. *Organization Science*, 12, 559-577.
- Wageman, R., & Mannix, E. A. (2005). *The uses and misuses of individual power in teams: Evidence from the field*. Manuscript submitted for publication.
- Wuensch, K. L. (2003). *Interrater agreement*. Retrieved March 24, 2005, from <http://core.ecu.edu/psych/wuensch/docs20/InterRater.doc>